SINGLE-MOLECULE BIOPHYSICS OF TOEHOLD-MEDIATED STRAND DISPLACEMENT

A Dissertation
Presented to
The Academic Faculty

By

Douglas William Bo Broadwater, Jr.

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology

August 2018

Copyright © Douglas William Bo Broadwater, Jr. 2018

SINGLE-MOLECULE BIOPHYSICS OF TOEHOLD-MEDIATED STRAND DISPLACEMENT

Approved by:

Professor Harold Kim, Advisor
School of Physics
Georgia Institute of Technology

Professor Kurt Wiesenfeld
School of Physics
Georgia Institute of Technology

Professor James C. Gumbart
School of Physics
Georgia Institute of Technology

Professor Peter Yunker
School of Physics
Georgia Institute of Technology

Professor Loren Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: July 19, 2018

I got power, poison, pain and joy inside my DNA.

I got hustle though, ambition flow inside my DNA.

Kendrick Lamar

To DaLin, with love

ACKNOWLEDGEMENTS

There is not enough space in this short remark to truly acknowledge the encouragement and support of every person who has helped me fulfill such a meaningful dream. My achievements are bolstered by a lifetime of inspiring moments from every caring person who pushed me to be better. From my youth, I want to acknowledge Mrs. Pamela Ray, Mrs. Lynda Wright, and Mr. Michael Williams for instilling me with a love of learning.

I want to thank the entire Mercer University Physics Department for introducing me to one of my deepest passions. Thank you to Dr. Matthew Marone for guiding me through my first research project. Thank you to Dr. Randall Peters for showing me there are many mysteries in physics at the mesoscale. Thank you to Dr. Sheng-Chiang Lee for introducing me to molecular biophysics in your Thermal Physics course. Thank you to Dr. Jose Balduz for enduring countless office hours of ridiculous what-if's and thought experiments. But, mostly, thank you for encouraging me to "do something in physics".

I owe a great debt of gratitude to the entire Harold Kim Lab. Thank you to Dr. Tung Le for his hard knocks style of training as it is truly the best way for a scientist to learn. Thank you to Dr. James Waters for Spaghetti Western Nights and for always having an answer. Thank you to Jiyoun Jeong for being the most patient labmate. Thank you to Gable Wadsworth for spitballing, car rides, and much needed venting sessions. I want to express my deep and sincere appreciation to my advisor, Dr. Harold D. Kim. As with most graduate students, my early understanding of doing science was narrow and stunted. Thank you for providing the tools and the space to make many mistakes.

Thank you to my supportive family who has been a wellspring of love and motivation. Thank you to my father and mother, Doug and Ronda Broadwater, for teaching me the value of hard work. Thank you to my wife and best friend, DaLin Broadwater, for sacrificing present comfort for future reward, for abandoning safety in exchange for adventure, and for chasing dreams with me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis focuses on three separate explorations into the nature of toehold-mediated strand displacement which is a reaction characterized by the swapping of one stably bound oligonucleotide for another. The first study seeks to improve current methodologies in single-molecule surface experiments using strand displacement. Presently, surfaces are prepared through time-consuming and costly cleaning and functionalization protocols which are only useful for a single trial of an experiment due to the irreversible binding of target molecules to the surface. To solve this problem, we develop a new method called ERASE which is powered by the inherent switching nature of strand displacement. Further, we showcase the wide applicability and adaptability of ERASE to different biophysical systems of interest.

Next, we use single-molecule fluorescence to study the effect of single mismatch position on strand displacement kinetics. We find a significant increase in the mean first passage time with mismatches proximal to the toehold. We highlight the 1D model's failure to account for the observed position dependence. Further, we show that addition of direct dissociation to the 1D branch migration model explains the observed data. To provide biophysical insight, we propose a simpler three-step model that explains mismatch-dependent behavior.

Finally, we explore the biophysics of the branch migration subprocess. We design FRET constructs to decouple the short-lived branch migration process from the overwhelmingly dominant toehold binding step. We further show a wide sequence-dependent diversity of waiting time distributions. We measure distribution changes between systems of "complementary invasion" which hint to dangling ends as the origin. Moreover, we design significantly less expensive single fluorophore constructs which allow for large scale sequence and length dependent studies.

CHAPTER 1
INTRODUCTION

There are few discoveries that have drastically changed our understanding of the universe like the discovery of the structure and function of deoxyribonucleic acid (DNA). Since Watson and Crick posited their double helix model in 1953 [1], the field of molecular biology has grown at a considerable rate and bridged the two foundational fields of genetics and biochemistry. As Francis Crick was a physicist by training, it is apparent that problems in molecular biology have interested physicists since its conception. Erwin Schrödinger beautifully summarizes a physicist's thoughts on biology by asking, "How can the events in space and time which take place within the spatial boundary of a living organism be accounted for by physics and chemistry?" [2]. Fueled by this question, scientists thought deeply and developed techniques to study biological molecules at an individual level giving rise to the field of single-molecule biophysics.

This thesis sprouts humbly from this enormous field to explore a particular biological reaction, toehold-mediated strand displacement. Toehold-mediated strand displacement occurs between three single strands of DNA when one unbound single strand invades and fully hybridizes a partially bound duplex. Strand invasion is critical to biology as it occurs in homologous recombination [3], a process that is both responsible for repair of double-strand breaks and for the source of genetic diversity via crossover events. Also, strand displacement has practical applications in nanotechnology as it is used in DNA construction and origami [4, 5, 6, 7, 8], biosensing [9, 10], DNA walkers [11, 12, 13, 14, 15], and DNA computation [16, 17, 18, 19, 20, 21]. Despite its fundamental role in biology and popularity in bioengineering, a physically intuitive and quantitative model of strand displacement does not exist for many scenarios. Further, the subprocess of two-way branch migration has never been explored with single-molecule approaches. Therefore, in this thesis, I will

discuss my work towards investigating, quantifying, and harnessing strand displacement at the single-molecule level.

I will begin this thesis with an introduction to nucleic acids and further motivate with toehold-mediated strand displacement including current applications and phenomenological models. I will then explore single-molecule experimental techniques used to study toehold-mediated strand displacement including fluorescence, microscopy, data acquisition, and analysis.

In Chapter 2, we will discuss our development of ERASE (Epitaxial Removal Aided by Strand Exchange), an application of toehold-mediated strand displacement which enhances surface-based single-molecule experiments. By exploiting the inherent switching behavior of strand displacement, we show that surfaces, which are time-consuming and costly to create, can be reset to an earlier state to allow for repeated experiments over the same field of view. We further show the wide applicability of our method as we extend it to DNA, RNA, and the full bacterial ribosome.

In Chapter 3, we explore the kinetics of toehold-mediated strand displacement and how it is affected by the presence of a single mismatch using single-molecule fluorescence. We find large changes depending on the mismatch position and present the effect of secondary structure on strand displacement kinetics. We compare our data to the current model and show how it must be adjusted to match our observations. Finally, we provide a simplified model that enhances biophysical understanding of the overall process.

In Chapter 4, we turn our attention to the subprocess of branch migration. We measure the first passage time distribution for several sequences using Fluorescence Resonance Energy Transfer (FRET) and find significant sequence dependence. We show a difference in first passage times for "complementary invasion" and suggest intermediate dangling ends as a source of differences as evidenced by the reversal of times with respect to RNA. We also increase the dangling end invader length to measure a slowdown of branch migration times that is consistent with current models.

In the final chapter, I will present some preliminary data from a different method to measure branch migration which uses a single fluorophore. This low-cost method opens the possibility to compare many different branch migration systems to provide a more complete biophysical understanding. We measure the branch migration time as a function of length of the displacement domain and confirm an expected increase in time with increasing length. Further, we measure the dependence of the waiting time distribution on the location of a single point mutation and discuss the implications.

1.1 A physicist's primer on nucleic acids

Traditionally, physics and biology exist on opposite ends of the natural sciences. Due to the large differences in training, knowledge-base, and culture, it can be difficult for physicists to begin tackling biological problems, much less ask interesting questions in the subject. In order to help bridge the knowledge-gap, this section serves as a very short introduction to DNA and RNA and their many exciting behaviors.

As we begin, it is helpful to define some terms that will appear often in our discussion and throughout the rest of the thesis. To the confusion of researchers new to the field, these terms will have many synonyms which is a result of the large number of people interested in studying nucleic acid biology. Indeed, the very lack of a general consensus for some terms is evidence of the fast growing origins of the field. The list below contains common terms, definitions, synonyms, and abbreviations.

Annealing

The binding and formation of hydrogen bonds to form double-stranded nucleic acid from individual strands. Synonyms: duplex formation, hybridization

Branch migration

A strand-exchange process which can occur between three or four complementary nucleic acid strands.

dsDNA

A molecule formed by two complementary single DNA strands. Synonym: duplex, molecule

Denaturation

The unwinding and separation of double-stranded DNA. Synonym: melting

Partial duplex

The molecule formed by annealing two strands of different length.

Strand

A polymer with nucleotide monomers covalently joined by phosphodiester bonds. Synonyms: molecule, Oligonucleotide (oligo), ssDNA, ssRNA

Toehold

Single-stranded region of a partial duplex.

Nucleic acids are polymers with nucleotides (monomeric subunits) built from three chemical components: phosphate, 5-carbon sugar, and a nitrogenous base. They can be largely separated into two catergories, DNA and RNA, depending on the absence or presence of the hydroxyl moiety on the $2'$ position of the pentose sugar. There are five nitrogenous bases which can be subdivided into purines, adenine (A) and guanine (G), or pyrimidines, cytosine (C), thymine (T), and uracil (U). Chargaff's rule [22] dictates the complementary Watson-Crick base pairing of G-C and A-T (U for RNA).

Double-stranded nucleic acids are formed through base pair interactions which are dominated energetically by base pairing (hydrogen bonds) and base stacking. GC base pairs share three hydrogen bonds and are, therefore, more energetically favorable compared to AT base pairs. As such, the GC content, which is the ratio of GC base pairs to the total number of base pairs, can be used a relative marker for the melting temperature of a duplex. Higher GC content correlates with higher melting temperature. Although base

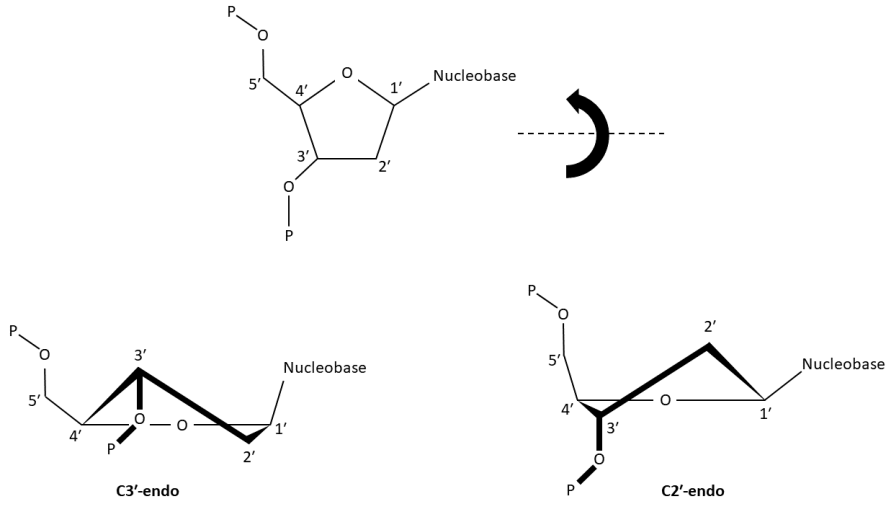


Figure 1.1: Endo sugar pucker conformations in DNA. The generic structure of a nucleotide is drawn so that it coincides with the plane of the drawing; however, a planar ring structure is not energetically favorable. If we rotate the molecule so that the 2′ and 3′ carbons are closest to us, we find two different conformations, C2′- and C3′-endo, where those carbons break the ring plane. Unlike DNA, RNA has a hydroxyl group on the 2′ carbon which can interfere sterically with the nucleobase thus discouraging RNA from adopting the C2′-endo pucker.

pairing is far more well known, base stacking is the dominant interaction. Base stacking plays a very important role in duplex formation and arises from the interaction of pi orbitals between neighboring rings. Stacking between individual bases is weak but base stacking is highly stabilizing for many bases.

A nucleic acid's ability to store a two-bit genetic code in its sequence is one of the most groundbreaking discoveries of the twentieth century. At first thought, it might seem that the sequence direction for nucleic acids is arbitrary; however, this is not true as nucleic acids are inherently polar due to the asymmetry of a single nucleotide. The molecule is extended through phosphate groups at the 5′ and 3′ carbon on the 5-carbon sugar. Sequences are defined to be read from the 5′ end to the 3′ end.

Finally, it is important to mention the three-dimensional shape of double-stranded nucleic acids. The double helix is a familiar concept from grade school. The principal features of the double helix include a right-handed sense, major and minor grooves, bases that are

perpendicular to the helical axis, $10.5\,\mathrm{bp}$ per helical turn, and $3.4\,\text{Å}$ of rise per base pair. However, this is one of three commonly known forms of double helix known as the B-form while the A-form and the Z-form are the other alternate geometries. While Z-form is left-handed and holds no definitive biological significance, A-form is important because it is the helical conformation of double-stranded RNA and RNA-DNA hybrids. This is due to the existence of the RNA 2′-hydoxyl group which stabilizes the pentose sugar in the C3′-endo sugar pucker over the C2′-endo sugar pucker. As DNA does not have the steric clash of the 2′-hydroxyl, it is free to adopt both sugar pucker conformations (see Figure 1.1). This has a stark mechanical effect as this causes dsDNA to be more flexible than dsRNA [23].

1.2 Toehold-mediated strand displacement

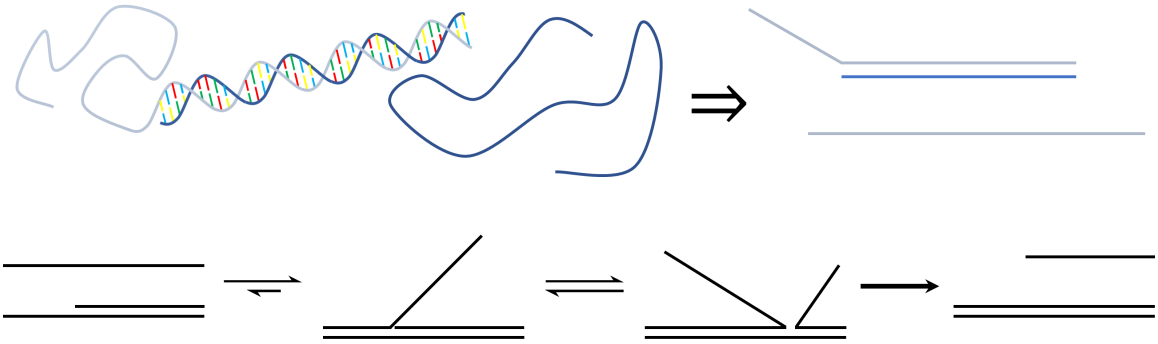


Figure 1.2: DNA abstraction and strand displacement. Above. A DNA partial duplex and single strand DNA are depicted. Details such as sequence and three-dimensional shape are abstracted away. Base pairing through hydrogen bonding is inferred by proximity to a neighboring DNA strand. Below. Strand displacement is a swapping reaction governed largely by a bimolelcular rate. The invader is the free single strand and initiates strand displacement as it binds to the single stranded region (toehold) on the substrate. The reaction advances through branch migration which is the consecutive exchange of homologous base pairs. This occurs until the incumbent is fully displaced by the invader. The dissociation of the incumbent is an irreversible step which intrinsically makes strand displacement an out-of-equilibrium process.

Toehold-mediated strand displacement is a reaction involving three single strands of DNA: invader, incumbent, and substrate (see Figure 1.2). In a typical reaction, the invader is fully or mostly complementary to the substrate. This is similarly true for the incumbent

except for the toehold region where the substrate remains single stranded. Strand displacement initiates when the single strand invader binds the toehold region of the partially bound duplex formed by the incumbent and the substrate. As this step is bimolecular, it is highly dependent on concentrations of the partial duplex and invader. Because of the reaction's high sensitivity to toehold binding, much work has been devoted to understanding the role of sequence [7] and length [24]. The reaction is thought to proceed through two-way branch migration where the invader and the incumbent compete for binding to the substrate [25]. The reaction is completed when the incumbent is fully dissociated. This step is thought to occur with dissociation of the base pair most distal to the toehold [25]; however, branch migration must always be in competition with direct dissociation as a possible pathway to completion. Obviously, there is no difference between direct dissociation and the final step of branch migration; however, the preceding steps must compete with direct dissociation of more than a singly bound base pair. Given that dissociation rates must decrease with increasing incumbent base pairs, it is not obvious that a single step of branch migration is always faster than direct dissociation.

Branch migration is thought to occur consecutively with individual homologous base pairs being exchanged one at a time [25]. If branch migration is naively modeled as a random walk, then the model for incumbent dissociation is similar to a classic problem in probability theory, the gambler's ruin [26]. Here, the competition between the incumbent and the invader for base pair binding on the substrate is similar to the poor gambler's attempts to win big against the house. Analysis of the gambler's ruin problem shows that for a fair game the gambler will always lose against a player with infinite wealth. An invader with a stably bound toehold is akin to the house with an infinite bankroll. Therefore, since the house always wins, it is similarly imagined that a stably bound toehold will always lead to successful strand displacement. This analogy breaks down for shorter toeholds as the probability of the invader dissociating before successful displacement increases; however, the fact that the invader will win remains. Fortunately, there is also a more physical argu-

ment to explain successful completion of strand displacement. By examining the initial and final states of strand displacement, it is clear that the invader-substrate duplex share more base pairs than the incumbent-substrate duplex. The increased base pairing implies a lower free energy. Therefore, thermodynamically, this reaction can proceed.

1.2.1 Nanotechnology applications of strand displacement

Toehold-mediated strand displacement has become a far-reaching and prevalent tool in the field of DNA nanotechnology. There are many great features that have attributed to its extensive interest and use in the field, including the wide availability of low-cost, bespoke oligos, DNA's overall suitability as a construction material with predictable structure and relatively straightforward thermodynamics, and the reaction's characteristic switching behavior with one clearly defined input and output.

The costs of reading and writing DNA have both steadily dropped over the past three decades [27]. Such fine-grained control over the sequence and length has allowed researchers to build impressive devices such as the DNA robot by Thubagere *et al.* that can sort molecular cargoes [28] (see Figure 1.3a). Powered by strand displacement, the robot sorted two different cargoes which were covalently attached to similar but unique DNA strands. Beyond sorting, DNA strand displacement is commonly used as an engine to power locomotion [29, 30]. Similar to the robot, DNA walkers function on a DNA origami surface where a walker molecule strand displaces with neighboring strands often at the expense a hairpin fuel source [31]. Dynamic behavior extends beyond physical motion as strand displacement has also been used to create biosensors. Recently, a DNA origami sensor was created to act as a nanoactuator [32]. The device was activated by successful strand displacement with a 22nt human miRNA, thereby acting as a detector.

Strand displacement also has found a myriad of applications in structural DNA nanotechnology. Andersen *et al.* used DNA strand displacement to serve as a 'key' to open a DNA origami box with functioning lid [4] (see Figure 1.3b). This holds exciting potential

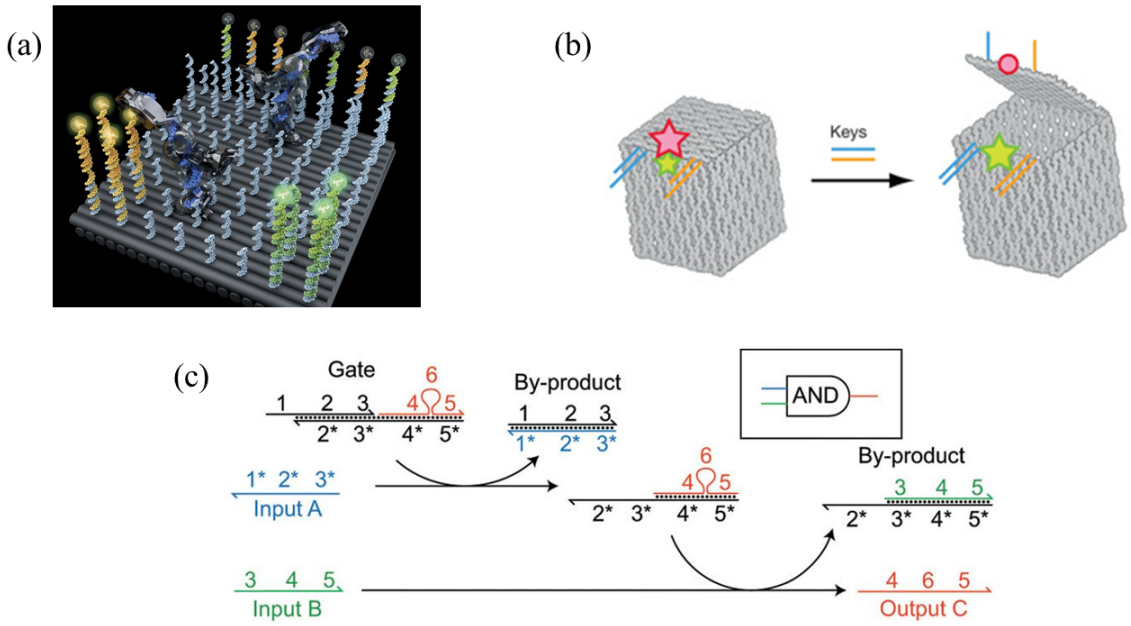


Figure 1.3: Strand displacement nanotechnology. (a) Cargo-sorting DNA robot. In this conceptual illustration, unsorted "cargo" is covalently attached to DNA strands and is arbitrarily sorted on a DNA origami raft (grey cylinders). Single and double DNA strands are arranged as vertical posts and can undergo strand displacement. Cargo strands travel in a 2D random walk across the raft until they are "sorted" into separate piles which correspond to their lowest free energy state. This figure is taken from reference [28]. (b) Nanoscale DNA box and controllable lid. A DNA origami box $42 \times 36 \times 36\,\mathrm{nm}^3$ with a lid "locked" by two partial duplexes. Two unique invader strands act as "keys" to control lid opening. A FRET pair is used to observe lid opening and closing. This figure is taken from reference [4]. (c) AND logic gate. Output C is only produced when both Input A and Input B are available to interact with the Gate. If either or both input strands are missing, then Output C is not produced. This matches the truth table for a logical AND gate. This figure is taken from reference [33]. These figures are shared with permission from the publishers American Association for the Advancement of Science (AAAS) and Springer Nature with the following license numbers: cargo sorting DNA robot - 4380561176563(AAAS), Origami box with lock and key - 4380570136773(Springer Nature), logical AND gate - 4380570709759(Springer Nature).

applications in nano-targeted drug delivery as it is not only controllable, but DNA origami methods have been explored as a solution to cross-resistance in the treatment of tuberculosis [34]. In an effort to push the scale of DNA construction, super-molecular structures which use DNA origami complexes as building blocks have demonstrated self-assembly [35]. In one instance, DNA strand displacement also has been used to disassemble self-assembling cages [36].

Besides being a highly useful structural material, DNA can also contain information. As such, DNA computation has found various functionality in DNA strand displacement where the abstraction as an input-output black box is the most useful characteristic to DNA computer engineers as it closely mimics a transistor switch. For example, Zhang *et al.* have demonstrated the processing power of DNA in creating logic gates through cascades of DNA strand displacement [33] (see Figure 1.3c). AND gates and logic inverters (NOT gates) [37] could conceivably be combined to create NAND gates. Further, as NAND gates are Boolean complete, all of classical computing can be recovered.

The one-to-one input/output relationship of strand displacement presents a potential concern as there is a lack of signal amplification; however, even this can be overcome through a strand displacement dependent amplification process [38] by stacking two separate incumbent strands in tandem on a single substrate. This allows a single invader strand to release both in a single reaction.

1.2.2 Strand exchange and branch migration in biology

The kinetics of strand displacement are dominated by the bimolecular, concentration-dependent toehold formation step. In most nanotechnology applications, the systems are perfectly complementary and the branch migration regions are shorter than $50\,\mathrm{bp}$. Further, the subsequent process of branch migration is very fast and, therefore, has little effect on the overall timescale of the reaction. As a result, bioengineers have paid little attention to the biophysics of branch migration as it is often thought of a symmetric 1D random walk

[25]. Nonetheless, as I shall show in Chapter 3, branch migration times are not modeled so simply. Moreover, branch migration is a critical component in three important biological systems: homologous recombination, the CRISPR/Cas system, and transcription (see Figure 1.4).

Homologous recombination is the process by which nucleotides are exchanged between a pair of complementary DNA sequences. This is used by the cell as a mechanism to repair breaks in the genome, as a source of genetic diversity in crossover events during meiosis I, and as a pathway for genetic exchange during horizontal gene transfer. As with most cellular processes, homologous recombination is protein-driven and, as such, can appear differently in different species. At the same time, however, it is highly conserved and exists in all domains of life [39]. Despite this rich diversity in behavior, homologous recombination bears a strong resemblance to branch migration. At the molecular level, homologous recombination must involve the exchange of base pairs between separate strands which, in the absence of proteins, is branch migration.

CRISPR/Cas (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated system) is the modern marvel of gene editing that is expected to revolutionize genetic engineering. It evolved as an adaptive immune system stored within the bacterial genome and used to recognize and destroy the genetic code of viruses [40]. Within the past decade, it has been recognized and demonstrated to have the ability to precisely edit locations in the genome. This is accomplished through a homology search of the Cas complex [41]. Further, mismatch experiments have hinted towards the possibility of behavior similar to one-dimensional branch migration [42].

Transcription is a key mechanism in the central dogma of molecular biology. It is the process by which mRNA is created from the DNA genetic library. Like most cellular functions, it is mediated through an ensemble of proteins whose most prominent member is RNA polymerase (RNAP) which unwinds double stranded DNA, guides nucleotides into place, and furthers elongation. Since the sole function of transcription is to harness nucleic

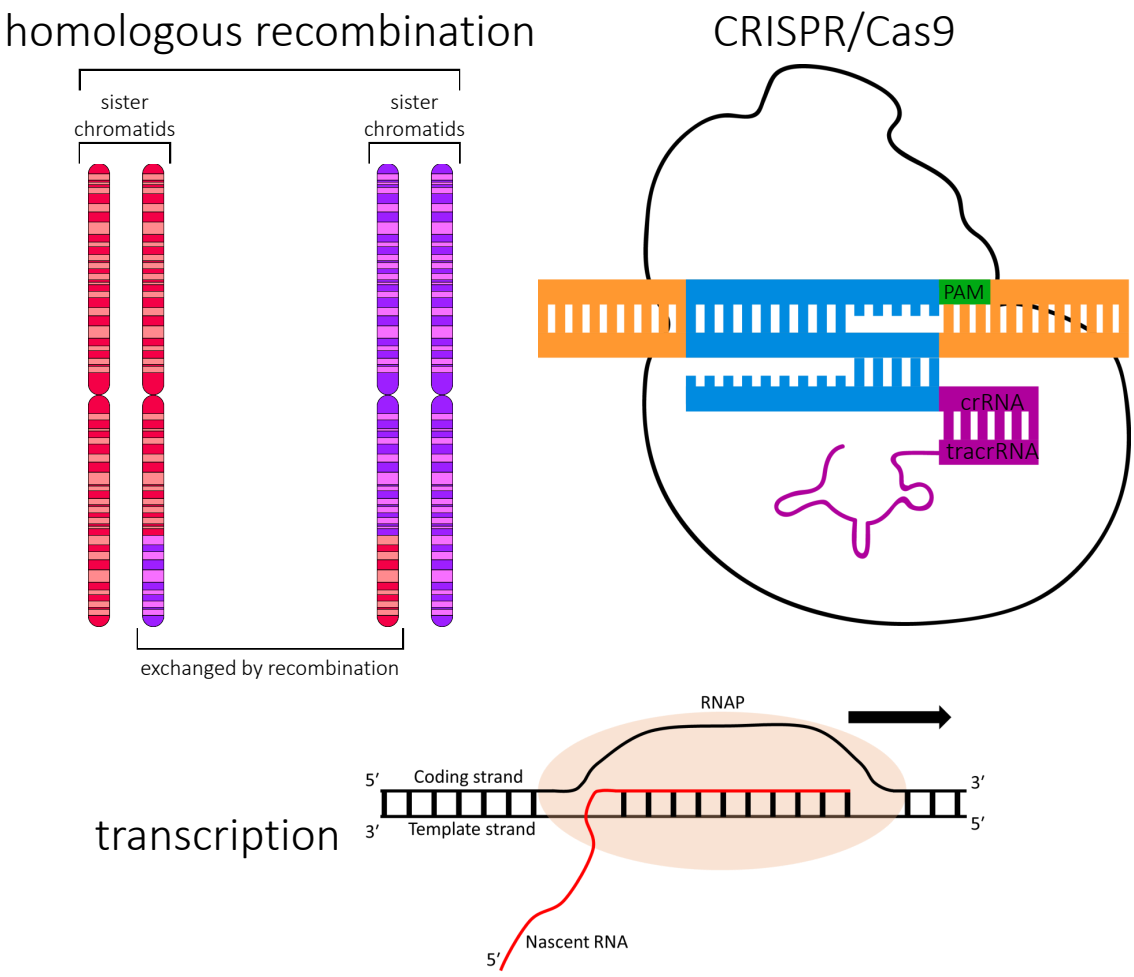


Figure 1.4: Strand displacement in biology. Homologous recombination is commonly introduced as the process responsible for genetic crossover events between homologous chromosomes; however, it is most widely used in cells to repair detrimental DNA breaks. At a molecular level, homologous recombination bears a striking resemblance to strand displacement as homologous strands exchange base pairs. CRISPR/Cas9 is the gene-editor's most prized tool as it allows for targeted and programmable gene editing. This is accomplished through a homology search for the protospacer adjacent motif (PAM) site followed by homologous exchange between the CRISPR RNA (crRNA) and the target gene. Finally, RNA polymerase transcribes single strand messenger RNA (mRNA) by reading the antisense (template) strand of DNA. The upstream strand of RNA remains bound and must be displaced by the sense (coding) strand.

acid's inherent base pairing ability to produce a complementary strand, a hybrid DNA-RNA double helix is a necessary product. The upstream edge of RNAP is very interesting as it has been shown that RNA-DNA and DNA-DNA base pairing strongly govern translocation and transcription rate [43]. Further, the mRNA must dissociate to fulfill its goal of being translated. Therefore, the sense DNA strand must invade and displace the RNA strand away from the antisense strand. A clear analogy exists with branch migration.

Given its relative simplicity compared to protein-mediated reactions, two-way branch migration could serve as a model system for more complicated biological processes. Deeper knowledge about the underlying biophysical mechanisms form a much needed foundation and could provide insight to systems that exchange nucleic acid base pairs.

1.2.3 Current understanding of strand displacement

Due to its compelling biological significance and ideal behavior as a nano-construction tool, DNA strand displacement has been intensively studied and modeled. For example, it is known that DNA strand displacement rate depends significantly on the toehold length [44]. It has previously been shown by Zhang *et al.* that the overall rate can span several orders of magnitude by varying the length of the toehold by only a few nucleotides, but this effect saturates for even the weakest toeholds that are longer than seven nucleotides in length [24]. They further introduced a phenomenological model that accurately describes the toehold-length dependent kinetic behavior. However, as a phenomenological model, it lacked any biophysical insight.

Srinivas *et al.* combined a 1D energy landscape model, a secondary structure kinetics model, and oxDNA [45], a coarse-grained 3D computational model, to answer some important biophysical questions about DNA strand displacement [25]. In this approach they were able to provide some biophysical explanations of the observed rate dependence on toehold length. They found that it could be explained with two factors: (i) a single step of branch migration is much slower than the fraying rate of an individual base pair

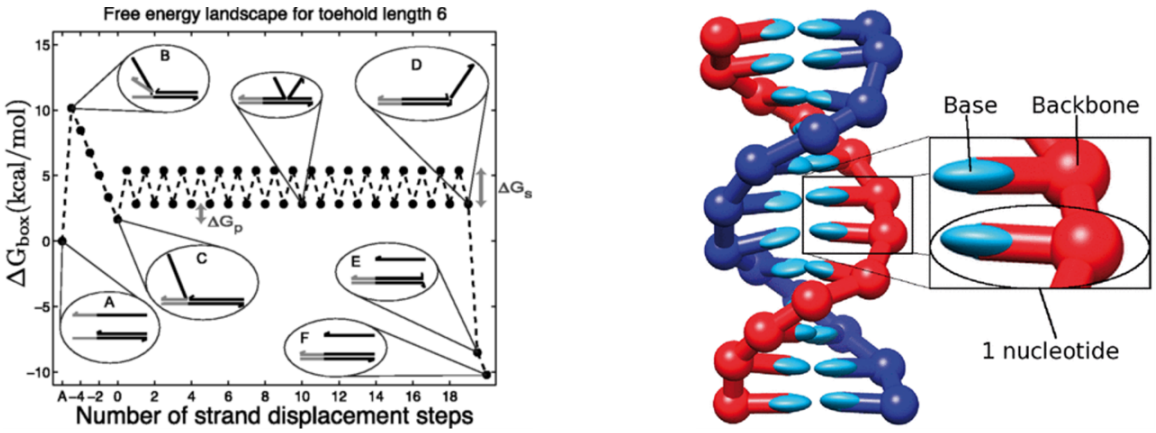


Figure 1.5: Strand displacement models. Left. A proposed 1D model for strand displacement maps the free energy for individual steps of strand displacement including two-way branch migration. This model assumes a few parameters including ΔG_s which is the height of the transition state for individual branch migration steps and ΔG_p which accounts for the entropic cost of creating the incumbent dangling end. Right. Pictured is a visualization of a DNA double helix from oxDNA, a coarse-grained nucleic acid simulation software which can be used to model strand displacement. Both of these figures were taken from reference [25] under the Creative Commons Attribution Non-Commercial License. More information is available from the journal's Open Access Inititive.

and (ii) the initial step of branch migration sustains a thermodynamic cost because it creates an additional overhang at the junction. They explain that the physical justification for the difference in speed of branch migration versus fraying is due to disruption of stacking interactions and structural rearrangement in branch migration. They further explain that the increased free energy barrier for the first step is most likely due to steric effects from creation of a dangling end in the incumbent.

The effect of a mismatch in DNA strand displacement was previously studied by Machinek *et al.* [46] They used bulk fluorescence methods to show that a base pair mismatch in the invader can provide kinetic control (beyond toehold length adjustment) over the rate of strand displacement. They also used oxDNA [45] to calculate the free energy landscapes

of strand displacement reactions containing a mismatch. They used these landscapes to qualitatively argue that the effect of mismatch dependence is a phenomenon largely due to kinetics as opposed to equilibrium thermodynamics. Finally, they also presented a beautiful experiment where they highlighted the differences in thermodynamic versus kinetic control of DNA strand displacement by allowing two invaders (one longer with a mismatch proximal to a toehold, another shorter one with a mismatch distal to the toehold) to compete in a toehold exchange experiment. For smaller timescales, the shorter invader with the distal mismatch quickly out-competed the longer invader with the proximal mismatch due to the kinetic mismatch effect; however, at longer timescales, the longer invader eventually displaced the shorter invader due to increased base pairing and, therefore, lower total free energy.

1.3 Unpacking the experimental toolkit for single-molecule biophysics

To better understand the cutting edge of biophysical research, it is helpful to summarize the overwhelming growth of biological knowledge and interest over the past century and a half. After Mendel's discovery of the fundamental law of inheritance [47], geneticists, such as Thomas Hunt Morgan, sought to understand life through the study of mutants which are organisms that have been augmented by the addition or subtraction of a small component. On the other hand, biochemists tackled the problem of life with an antipodal approach whereby components from organisms are studied individually. These philosophies complement each other and are well suited to understanding biological function. As an illustration, imagine wanting to understand the function of how yeast cells digest a particular sugar. The geneticist would isolate strains of yeast that have lost the ability to process the sugar and look to identify the missing component. Alternatively, the biochemist would digest and fractionate yeast cells until they found the enzyme that digested the sugar. Despite the high degree of complementarity, the two fields had surprisingly little interaction until the discovery of the structure of DNA which opened the doors to the new field of molecular biology [48].

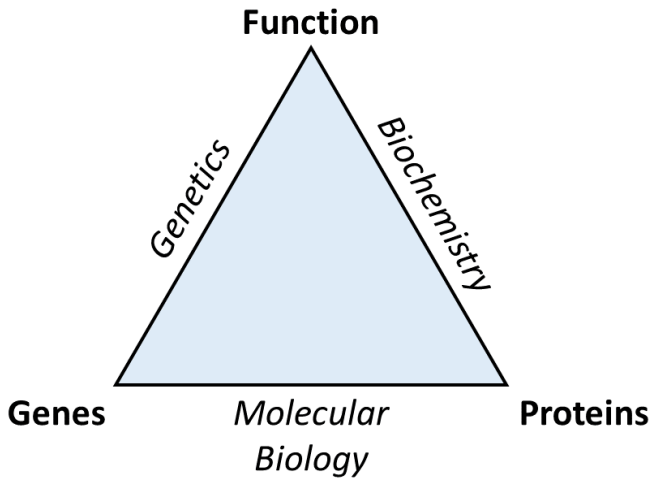


Figure 1.6: Roadmap of Modern Biology [48]. The study of biological function is addressed by the two highly complementary fields of genetics and biochemistry. While genetics observes behavior by changing one gene at a time, biochemistry fractionates cellular proteins to study individual behavior. As the picture suggests, molecular biology spans the two fields and studies the interaction of genes and proteins.

While geneticists and biochemists are interested in studying the relationship between biological function with genes and proteins (respectively), molecular biology bridges the two and seeks to understand the relationship between genes and proteins (see Figure 1.6). In the past 70 years, field-defining techniques such as gel electrophoresis, polymerase chain reaction, and molecular cloning have ushered a new era of biological understanding; however, these techniques only allow measurements on an ensemble or bulk-collection of molecules which effectively limits quantitative information to average characteristics. The true reductionist dream was realized with the observation of individual ion channels using the patch clamp technique [49]. Since then, large advances in biological understanding have developed due to the invention of powerful techniques like single-molecule fluorescence and atomic force microscopy. It is remarkable that a bulk of biophysical understanding is a result of either pushing/pulling molecules [50] or optically inferring molecular motion [51]. The former is accomplished through well-known techniques such as optical

tweezers, magnetic tweezers, and atomic force microscopy while the latter is accomplished through fluorescence-based techniques and tethered particle motion.

Biophysical force probes have uncovered detailed mechanisms in a wide range of systems including DNA compaction [52], DNA helicase kinetics [53], and Cascade-induced R-loop formation [54]. Such techniques are capable of probing dynamic behavior of nucleic acids. Indeed, tweezers are commonly used to unzip [55] and stretch [56] DNA. Given the high similarity between strand displacement and DNA dissociation, it is conceivable that these techniques could provide mechanistic insight; however, this is unlikely due to the small length scale of interest for strand displacement ($\sim 15\,\mathrm{nt}$) which is well below the working scale of more than $1000\,\mathrm{nt}$ for typical tweezers experiments. Further, strand displacement is an inherently non-equilibrium process which, unlike equilibrium processes, can only provide one transition/reaction per molecule. This significantly decreases the throughput for force probe techniques and renders them unfeasible.

Conversely, fluorescence-based techniques overcome these issues. The working length scale of FRET is $2\,\mathrm{nm}$ to $8\,\mathrm{nm}$ [57] which can probe the switch-like behavior of strand displacement through one or two fluorescent labels on separate DNA strands. Moreover, throughput is maintained at pragmatic levels by imaging a field of view containing hundreds of distinct strand displacement reactions. Unfortunately, there is a significant limitation to fluorescence imaging methods: time resolution. At a reasonable throughput, the fastest time resolution is on the order of $\sim 4\,\mathrm{ms}$. Faster methods exist such as the single molecule field effect transistor (smFET) which works by attaching nucleic acid molecules to a single defect in a carbon nanotube and monitoring electronic signals based on intrinsic molecular charge [58]. smFET can observe DNA hybridization kinetics with $\sim 1\,\mu\mathrm{s}$ resolution; however, the concern of throughput remains an issue with this method as well.

The experimental work in this thesis relies heavily on the ubiquitous techniques of Fluorescence Resonance Energy Transfer (FRET) and Total Internal Reflection Fluorescence (TIRF) microscopy. In the following subsections, I will explain a few bare essentials in the

single-molecule biophysicist's toolkit including the principles of fluorescence and FRET, how TIRF microscopy is implemented to achieve a high signal-to-noise ratio, and how quartz slides and glass cover slips are transformed into arenas upon which the experiments are conducted.

1.3.1 Fluorescence and FRET

Since molecules exist at a length scale far below the diffraction limit, clever techniques must be developed to visualize them. A rather direct method of observation is to covalently attach a bright fluorophore to the molecule of interest. This is similar to blasting a bright spotlight or beacon to some observer out in space. We can understand the photophysics of such processes through a very simple model of absorption and emission (photoluminescence).

$$\text{Excitation: } E_0 + \hbar\omega_{ex} = E_1 \tag{1.1}$$

$$\text{Emission: } E_1 = E_0 + \hbar\omega_{em} + \text{heat} \tag{1.2}$$

In this model, the molecule is raised from a lower energy state (E_0) to a higher energy state (E_1) by absorbing a photon with energy $\hbar\omega_{ex}$. The system decays to a lower energy state and emits a photon with energy $\hbar\omega_{em}$. In a phenomenon known as Stokes shift, the emission photon will often have a lower energy than the excitation photon, thus the excess energy is given off as heat. The excitation and emission spectra of Cy5, a commonly used dye, is shown in Figure 1.7. The Stoke shift, properly defined as the difference in wavelength between the spectral maxima, is clearly visible in the plot. Photoluminescence can be broadly divided into two categories of emission: fluorescence and phosphorescence. The differences between the two mechanisms are best described by a Jablonksi diagram (see Figure 1.8). Electronic energy levels are represented by bold horizontal lines while thin horizontal lines represent energy levels for excited vibrational states. Fluorescence oc-

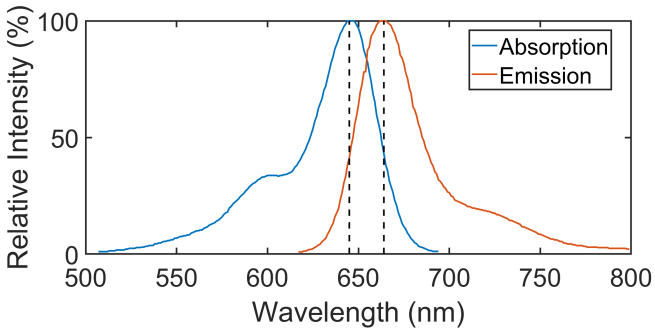


Figure 1.7: Cy5 spectra and Stokes shift. The relative intensity for excitation and emission of Cy5 are plotted against the wavelength. The maximum of each spectrum is denoted by a dashed line. The Stokes shift can be observed as the difference in wavelength between these two peaks indicating an energy difference between the absorbed and emitted photon. Data was taken from ThermoFisher Scientific [59].

curs when the system transitions from an excited electronic singlet state to a lower energy singlet state while emitting a photon with energy equal to the difference in energy levels. Such processes occur with characteristic timescales of $10^{-9}\,\mathrm{s}$ to $10^{-8}\,\mathrm{s}$ [60]. Phosphorescence occurs when the system relaxes through intersystem crossing. In such a process, the spin of the excited electron is reversed through spin-orbit coupling as it enters a triplet state. Phosphorescence has a much slower characteristic time of $10^{-3}\,\mathrm{s}$ to $1\,\mathrm{s}$ (or longer) [60] due to its spin-flip transition from the excited triplet state to the ground singlet state. Incidentally, these differences in time scales explain why, upon flicking the light switch, fluorescent bulbs no longer shine while phosphorescent glow-in-the-dark toys do. When using fluorescence in the lab, two practical issues will be encountered: photoblinking and photobleaching. Photoblinking is defined as a temporary (but repeated) drop in fluorescent intensity while photobleaching is an irreversible phenomenon where the fluorophore is said to "switch off". Clearly, both can wreak havoc in an experiment. For example, imagine monitoring the binding of a labeled transcription factor to DNA. It is impossible to distinguish binding and unbinding from an entirely bound protein with a photoblinking fluorophore. Similarly, a photobleached fluorphore can no longer act like a spotlight to provide information. Clearly, this poses a serious issue and must be dealt with. Photoblinking is

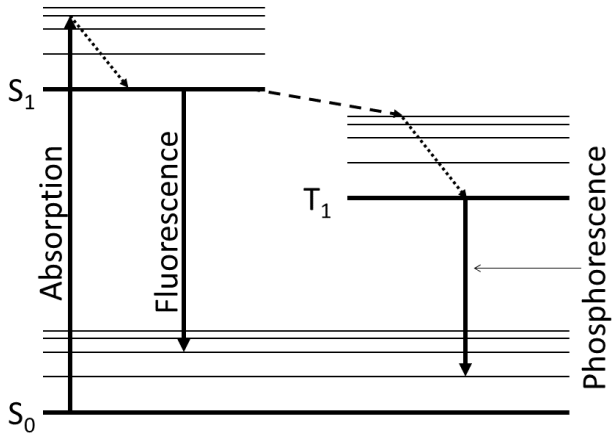


Figure 1.8: Jablonski diagram of photoluminescence. This schematic illustrates the electronic states of a photoluminescent molecule. The molecule absorbs a photon and is excited from the ground state of S_0 to an excited vibrational state of S_1, another singlet state. After non-radiative transfer (dotted line), the system can decay to an excited vibrational state of S_0 via fluorescence. Alternatively, the molecule can undergo intersystem crossing (dashed line) and transition to a triplet state, T_1, through a spin flip. Due to the"forbidden" spin transition, the system will take longer to decay back to S_0 via phosphorescence.

believed to be related to the formation of long-lived non-fluorescent triplet states as a result of non-radiative transitions [61]. Molecular oxygen is an effective quencher of the triplet state and thus is a potential choice for inhibiting photoblinking. However, in an ironic fate, molecular oxygen is a primary driver of photobleaching [62]. This leaves the experimenter with a choice: retain molecular oxygen and find a way to fix photobleaching or remove molecular oxygen and find a way to fix photoblinking. Common practice has firmly chosen the latter with 2-mercaptoethanol or Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid) being popular alternatives to singlet oxygen as a triplet state quencher. Enzymatic oxygen-scavenging systems such as glucose oxidase and catalase (GODCAT) and protocatechuic acid (PCA)/protocatechuate-3,4-dioxygenase (PCD) remove oxygen through enzymatic catalysis thereby reducing photobleaching. While photobleaching is never fully halted, the timescale of photobleaching can usually be decoupled from the timescale of experimental interest.

A single fluorophore label on a molecule can inform us about the existence and spatiotemporal position of a molecule. In the context of binding assays, this can be extremely informative. We have even designed and studied strand displacement with a single fluorophore construct. The results of which are contained in Chapter 5. At first glance, it seems that the extent of knowledge we can gain from such experiments is limited to simple kinetic binding assays albeit with the addition of spectrally distinct fluorophores to distinguish multiple molecular species. However, the phenomenon of FRET opens a wide door of possibilities such as protein folding [63], conformational changes of CRISPR/Cas9 [64], and DNA bending [65]. Details of intermolecular conformational changes can be tracked by using two dyes which act as proximity sensors. This is accomplished by energy transfer from an excited fluorescent dye called a donor to a nearby unexcited fluorescent dye called an acceptor through a dipole-dipole interaction.

FRET is inherently quantum mechanical since it involves a discrete electromagnetic interaction (photon) between two molecules. Further, it occurs at lengths much smaller than the photon wavelength which corresponds to the classical near-field region. Finally, the photon exchanged between the donor and acceptor is short lived and undetectable (virtual) as it violates the conservation of energy and momentum. The involvement of virtual photons implies a fully quantitative understanding that must come from quantum electrodynamics. While a complete theoretical treatment is beyond the scope of this thesis, I will provide a heuristic derivation which follows from reference [66] that starts with Fermi's golden rule and the classical dipole-dipole interaction energy to arrive at a famous result. We begin with a description of the process.

$$D + \hbar\omega_{ex} \rightarrow D^* \tag{1.3}$$

$$D^* + A \rightarrow D + A^* \tag{1.4}$$

$$A^* \rightarrow A + \hbar\omega_{em} \tag{1.5}$$

This is similar to standard photoluminescence with the additional step of energy transfer from an excited donor, D^*, to an unexcited acceptor, A. Fermi's golden rule describes the transition rate between energy eigenstates in a continuum.

$$k_t = \frac{2\pi}{\hbar}\left|\left\langle \Psi_f \left| \hat{H} \right| \Psi_i \right\rangle\right|^2 \rho \tag{1.6}$$

The initial wavefunction is $\Psi_i = \psi_{D^*}\psi_A$ and the final wavefunction is $\Psi_f = \psi_D\psi_{A^*}$. The classical dipole-dipole interaction energy is the product of dipole moments, μ, so that

$$E = \frac{1}{4\pi\epsilon_0}\frac{\kappa\mu_D\mu_A}{r^3} \tag{1.7}$$

where the orientation factor, κ, is defined as $\kappa = \hat{\mu_D} \cdot \hat{\mu_A} - 3(\hat{\mu_D} \cdot \hat{r})(\hat{\mu_A} \cdot \hat{r})$. A quantum mechanical treatment requires the dipole moment to be the transition dipole moment

$$\mu_X = \int \psi_X^* e\vec{r}\psi_{X^*} d^3r \tag{1.8}$$

where ψ_X^* is the complex conjugate of the ground state wavefunction, ψ_{X^*} is the excited state wavefunction, and e is the electron charge. The density of states can be connected to the normalized spectral overlap integral

$$J = \frac{\int f_D(\lambda)\epsilon_A(\lambda)\lambda^4 d\lambda}{\int f_D(\lambda) d\lambda} \tag{1.9}$$

where f_D is the donor emission spectrum and ϵ_A is the molar extinction coefficient of the acceptor. Putting this all together yields

$$k_t = \frac{9000(\ln 10)\kappa^2 Q_D J}{128\pi^5 n^4 N_A \tau_D r^6} \tag{1.10}$$

where Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor, n is the index of refraction, N_A is Avogadro's number, and τ_D is the mean lifetime of the

excited donor state D^*. It is helpful to consider the energy transfer efficiency simply known as the FRET efficiency

$$E = \frac{k_t}{k_t + \tau_D^{-1}} = \frac{1}{1 + (\frac{R_0}{r})^6} = \frac{I_A}{I_D + I_A} \tag{1.11}$$

where the Förster radius is defined as $R_0 = r(\tau_D k_t)^{1/6}$, I_A is the intensity of acceptor, and I_D is the intensity of the donor. FRET efficiency defines the relationship between distance and experimentally measurable intensities. The ability to infer distances from experimentally measured fluorophore intensities is what makes FRET such a powerful single-molecule tool as nanoscale ruler.

There are many choices of FRET pairs available. One of the most conventional FRET dye pairs are Cyanine 3 (Cy3) and Cyanine 5 (Cy5). Their wide utility is justified by several characteristics, including small size ($\sim 10\,\text{Å}$), high brightness (high extinction coefficient and quantum yield), and well-known labeling protocols. Most importantly, their spectra overlap significantly which is required for FRET (see Equation 1.9). The absorption and emission maxima are $550\,\text{nm}$ and $565\,\text{nm}$ for Cy3 and $655\,\text{nm}$ and $667\,\text{nm}$ for Cy5 [57]. Their difference in spectral emissions is also important as it allows for spectral separation through filtering. Cyanine dyes form a closely related family of chemical structures with differences mostly existing in the number of conjugated pi bonds between the ring structures. This unique structure allows for a special application of the simplest quantum mechanical model, the particle-in-a-box. Figure 1.9 justifies the coarse model. The well known solution of the energy eigenvalues dictates the expected difference between energy levels to depend on the length, $\Delta E = \frac{\pi^2\hbar^2}{2mL^2}(n_f^2 - n_i^2)$. Therefore, the wavelength λ_{abs} of an absorbed photon is proportional to the square of the length of the well, $\lambda_{abs} \sim L^2$. Such a crude model provides an outstanding agreement with experiment; however, this has been shown to be less related to the flat energy landscape assumption and more related to the accidental fact that the electronic energy difference (HOMO-LUMO gap) is not perturbed

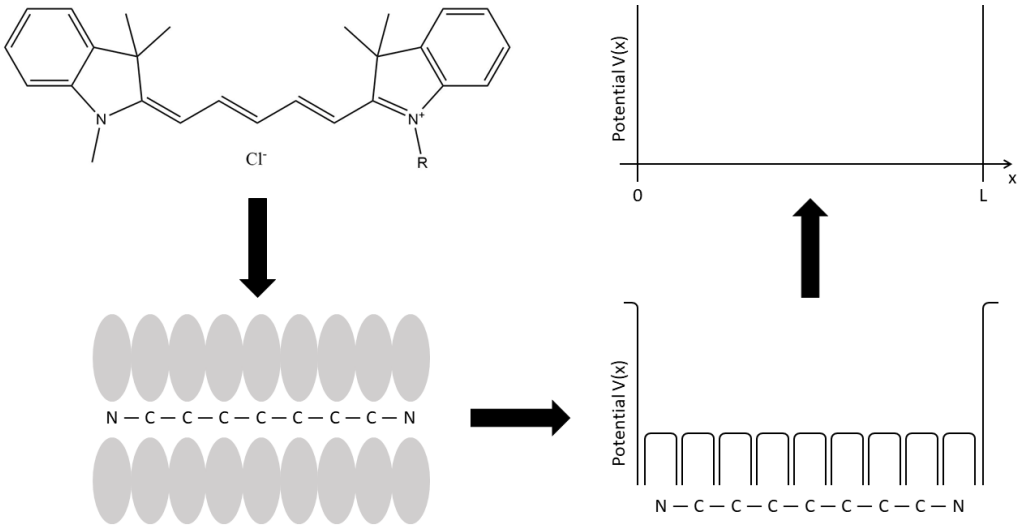


Figure 1.9: Particle-in-a-box model for cyanine dyes. The chemical structure of Cy5 is shown in the upper left. In the bottom left, the conjugated pi bonds create a bridge for electrons that do not belong to a single atom but are shared between a group of atoms. In the bottom right, a coarse model can be intuited where the nitrogen atoms serve as barriers and the conjugated pi bonds provide a relatively flat energy landscape. In the top right, the model is naturally extended to the analytically solvable particle in a box. This figure is adapted from reference [67] under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 United States License.

by the periodicity in the potential. [68].

1.3.2 Total Internal Reflection Fluorescence Microscopy (TIRFM)

A firm understanding of the principles of microscopy is absolutely essential to observe single fluorescent molecules. In fact, the photons leaving the fluorescent label are completely useless to the experimenter without a controlled method to excite the fluorophore then capture and record the resulting emission. For the most part, microscope construction is guided by an understanding of optical filters, simple geometric optics, and the thin lens equation, $1/o + 1/i = 1/f$, where o and i are the distances to the lens from the object and the image, respectively, and f is the focal length of the lens; however, a complete appreciation for the physical details of total internal reflection (TIR) requires physical optics.

We begin with a physical description of electromagnetic plane waves encountering a boundary formed between two homogeneous media. Any arbitrarily polarized plane wave can be decomposed into 2 orthogonally polarized plane waves. The conventional choices are the transverse electric (TE) and transverse magnetic (TM) polarizations so named for the corresponding field that is perpendicular to the plane of incidence (see Figure 1.10). These also respectively correspond to their confusing but canonical names: s and p. All electromagnetic phenomena must satisfy the continuity equations across a boundary (defined by normal vector $\hat{\mathbf{n}}$ which points along the positive z direction):

$$\hat{\mathbf{n}} \cdot [\mathbf{D}_1 - \mathbf{D}_2] = \hat{\mathbf{n}} \cdot [\mathbf{B}_1 - \mathbf{B}_2] = \hat{\mathbf{n}} \times [\mathbf{E}_1 - \mathbf{E}_2] = \hat{\mathbf{n}} \times [\mathbf{H}_1 - \mathbf{H}_2] = 0. \tag{1.12}$$

If we define the plane of incidence as the x-z plane, these conditions can only be met at the boundary if all phases are equal which implies $\omega_I = \omega_R = \omega_T = \omega$ and $k_{Ix} = k_{Rx} = k_{Tx}$. Since the incident and reflected wave occupy the same medium, they must have equal magnitudes, $k_I = k_R = k_1$. Equal magnitudes and equal x components of the incident and

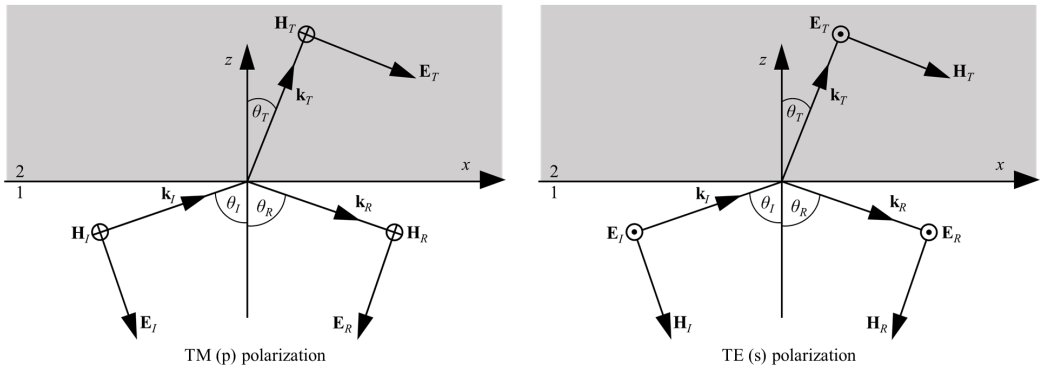


Figure 1.10: Reflection and refraction of plane waves. The wave vectors ($\mathbf{k}$) and field vectors ($\mathbf{E}$ and $\mathbf{H}$) of the incident, reflected, and transmitted plane waves are shown for two different orthogonal polarizations. In this figure, the index of refraction of the first medium is smaller than second medium hence θ_T is smaller than θ_I. Circles and crosses represent vectors coming out of and into the incidence plane. This figure is redrawn from Figure 17.4 in reference [69] with permission from Cambridge University Press. More information is available at `http://www.cambridge.org/about-us/rights-permissions/permissions/`.

reflected wave vectors imply the law of reflection.

$$\theta_I = \theta_R = \theta_1 \tag{1.13}$$

For symmetry, we write $\theta_T = \theta_2$. Comparing the incident and transmitted wave vectors, we derive Snell's law of refraction.

$$k_{Ix} = n_1 \frac{\omega}{c} \sin\theta_1 = k_{Tx} = n_2 \frac{\omega}{c} \sin\theta_2 \implies n_1 \sin\theta_1 = n_2 \sin\theta_2 \tag{1.14}$$

When light passes from a higher to a lower index of refraction, Snell's law requires the existence of a critical angle $\theta_c = \sin^{-1}(n_2/n_1)$ at which $\theta_2 = \pi/2$ and all light is reflected. This is the phenomenon of TIR. If we examine the z-component of the transmitted wave

vectors, we find an amazing result.

$$k_{Tz} = \sqrt{k_T^2 \cos^2\theta_2} = \frac{\omega}{c}\sqrt{n_2^2(1-\sin^2\theta_2)} = \frac{\omega}{c}\sqrt{n_2^2 - n_1^2\sin^2\theta_1} = i\frac{\omega}{c}n_1\sqrt{\sin^2\theta_1 - \sin^2\theta_c} \tag{1.15}$$

The last expression shows that k_{Tz} is purely imaginary when $\theta_1 > \theta_c$ (i.e. TIR). Upon closer inspection of the last term, we can identify the characteristic length scale as the penetration depth $d = \frac{\lambda_0}{2\pi n_1\sqrt{\sin^2\theta_1 - \sin^2\theta_c}}$ where λ_0 is the wavelength of the light in vacuum. Because the solution to a plane wave can be written as an oscillating exponential (imaginary argument), we can now see that the electric field in the lower index medium exponentially decays in the z direction.

$$\mathbf{E}_2(x,z) = \mathbf{E}_T e^{-z/d}\exp[i(k_1 x\sin\theta_1 - \omega t)] \tag{1.16}$$

The exponentially decaying refracted wave is known as an evanescent wave. For visible wavelengths and refractive indices close to water (~ 1.33), the penetration depth takes on values between $100\,\mathrm{nm}$ and $200\,\mathrm{nm}$ [70, 71]. This shallow penetration depth is the key to using TIR to study single molecules. By providing excitation close to the surface, we can observe individual molecules that are nearby (or bound to) the surface separate from the bulk of molecules that may be freely diffusing in solution.

Generally, there are two configurations through which to accomplish TIRFM: objective-type TIRFM (cis-geometry) where the excitation and emission light paths are both focused by the objective lens and prism-type TIRFM (trans-geometry) where the excitation beam passes through a quartz prism sitting above the sample. Prism-type TIRFM can often pass the excitation beam far above the critical angle to dramatically decrease the penetration depth; however, the emission signal must pass through the entire specimen which can cause spherical aberrations for thick samples. Further, fluorescence is anisotropic [73] and stronger toward glass which means prism-type has a lower collection efficiency than

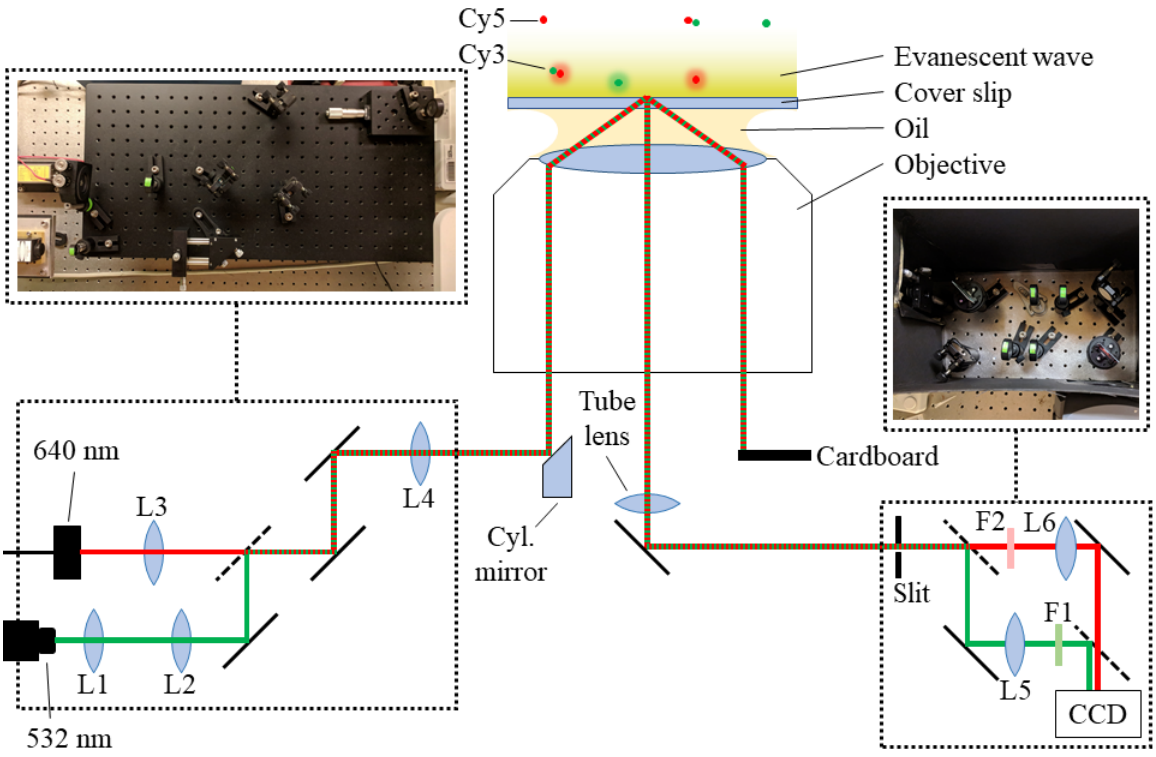


Figure 1.11: Objective-type TIRFM. Excitation begins on the left with 640 nm and 532 nm laser sources. Light from the 532 nm laser beam is expanded as it passes through L1 which acts as a diverging lens and is collimated by L2. Meanwhile, light from the 640 nm laser exits a fiber optic cable and is collimated by L3. The light passes through two pairs of mirrors (broadband mirrors are shown in solid lines, dichroic mirrors are represented by dashed lines). The number of mirrors are important because a pair of mirrors can place any incoming laser beam into any desired orientation. Thus, the first pair of mirrors is used to "walk the beam" of the 532 nm laser to coincide with the 640 nm laser, The second pair of mirrors allows both laser beams to pass through L4, reflect off the small cylindrical mirror [72], and focus onto the back focal plane of the objective (not shown). The beam passes through index-matching immersion oil and the cover slip then totally internally reflects at the specimen interface producing an evanescent wave. After TIR, the excitation light is blocked by a small piece of blackout cardboard. The emission from the sample is collimated by the objective into the infinity space and exits the microscope. An adjustable slit is placed at the focal point of the tube lens to crop the image for later placement on the electron-multiplying charge-coupled device (EMCCD). The emission passes through a custom-made beam splitter to separately visualize the Cy3 and Cy5 spectra. Relay lenses (L5 and L6) form the image on the EMCCD with 1:1 magnification. Background signal is removed with a bandpass filter (F1) and longpass filter (F2).

objective-type [74]. By contrast, objective-type TIRFM does not require a quartz prism and, therefore, allows access to the top of the slide. While objective-type has a higher background due to the laser passing through oil-immersion optics, it can be accomplished with glass slides which are cheaper than the quartz slides required for prism-type. Moreover, in prism-type TIRFM, molecules must be immobilized on quartz slides which have to be cleaned thoroughly before experiment, whereas objective-type TIRFM employs single-use glass cover slips which are much cheaper.

For these reasons, we employed objective-type TIRF to visualize fluorescently labelled molecules attached to a glass cover slip (see Figure 1.10). Conventionally, objective-type TIRF is accomplished with a multi-band dichroic mirror to reflect excitation laser light and allow emission signal to pass through; however, we utilize a small cylindrical mirror that direct the laser light towards the objective [72] and a small piece of optical blackout board to block the light from the excitation pathway. The mirror and cardboard are mounted onto stages that allow for adjustment of the beam path through the objective. As the beam is moved toward the edge of the objective, the beam exits the objective at a steeper angle. Once the critical angle is reached, the laser beam totally internally reflects at the boundary and is collected by the objective. As discussed before, an exponentially decaying field with identical wavelength is created on the other side of the boundary [69]. This evanescent wave decays exponentially with a characteristic length of around $200\,\mathrm{nm}$ [70, 71] which can be used to excite a fluorophore near the surface. The excitation pathway, which could either be $532\,\mathrm{nm}$ or $640\,\mathrm{nm}$ laser light, includes a telescope for beam expansion, a lens for focusing on the back focal plane of the objective, and two pairs of mirrors which allow full control of the alignment of both lasers. The emission light passes through a slit for cropping, a beam splitter for spectral separation, and is filtered for background signal and focused onto an EMCCD for recording.

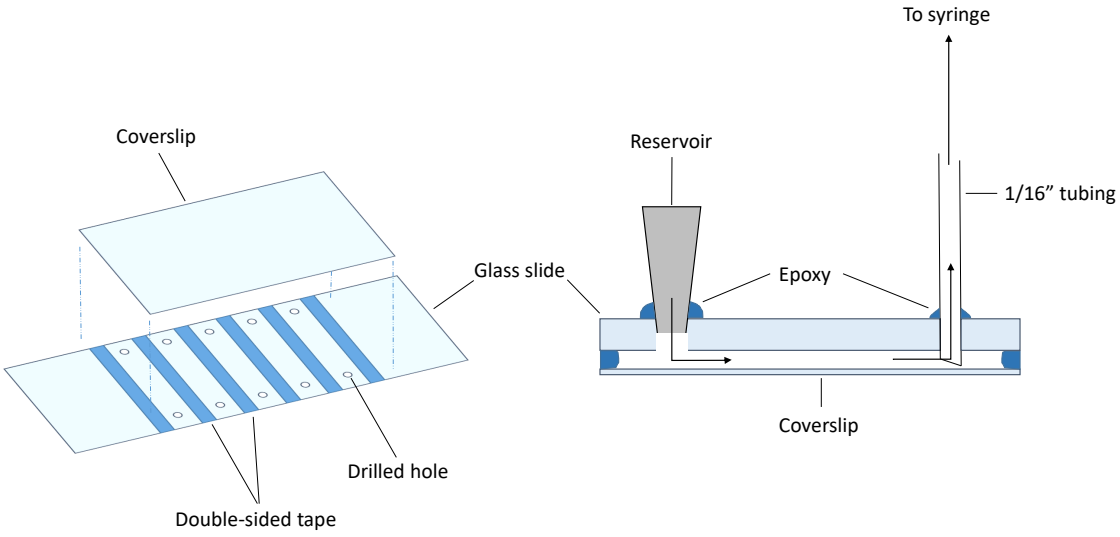
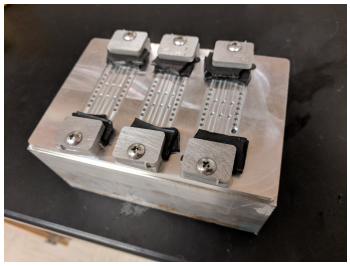


Figure 1.12: Flow cell design. On the left is an exploded, isometric view of the flow cell. The double sided tape is cut to form individual reaction channels so that many experiments can be performed on each flow cell. On the right is a side view of a complete flow cell. 1/16” tubing is inserted into the drilled holes so that a pump can withdraw buffer and sample from a reservoir (constructed from a micropipette tip) through the flow cell and into a syringe. 5-minute epoxy is used to seal any gaps left from construction.

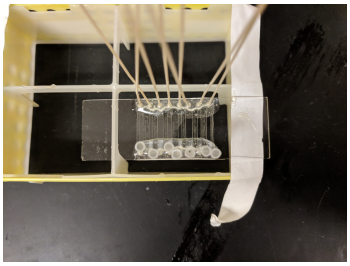
1.3.3 Setting the experimental stage

After discussing fluorophores as well as how to illuminate and image them with TIRFM, one final tool remains to be explored: mise en scène of the experimental arena, including the surface attachment of molecules and construction of flow cells. In an effort to increase throughput of experimental data, flow cells were constructed with perforated glass slides, cover slips, double sided tape, and 5-minute epoxy (Fig. 1.14). The strips of double sided tape serve as $0.1\,\mathrm{mm}$ tall channels guides. Care must be taken to ensure the tape has properly sealed to both the slide and cover slip or else risk leaking buffer into other channels thereby cross-contaminating while also losing an airtight seal. If precise control of flow is critical to the experiment, then plumbing can be added that will attach to a syringe pump. Pipette tips serve as excellent reservoirs for buffer to be introduced to the system while 1/16” tubing can be epoxied to the exit port and relayed to the syringe.



(a) CNC jig

(b) Plasma cleaner



(c) Flow cell with plumbing

Figure 1.13: Constructing flow cells. We designed and machined an aluminum jig for the precise and rapid drilling of holes into glass slides. A CNC machine was used to machine the jig and to drill holes in slides. Plasma etching cleans glass slides and cover slips by bombarding the organic debris with highly energized nitrogen plasma. Clean slides and cover slips can be assembled into flow cells with double sided tape and epoxy.

Holes can be drilled into glass with 0.05” diameter diamond-plated drill bits. Holes can be manually drilled by reconfiguring a handheld rotary Dremel into a miniature drill press; however, manually drilling holes in glass slides is difficult and time-consuming due to the high level of precision and care required. To overcome this, we designed an aluminum jig (see Figure 1.13a) that can firmly hold three slides and is to be used in conjunction with a Computer Numerical Control (CNC) machine that will precisely and accurately drill holes. Automation of the hole-drilling process tremendously increases throughput and reproducibility in the glass slides.

Prior to construction, slides and cover slips must be cleaned to remove unwanted fluorescent signal and to isolate the system. This can be accomplished through either chemical or plasma etching. Chemical etching involves several rounds of sonication in deionized water, acetone, and potassium hydroxide. This method is costly and wasteful as fresh chemicals must be used each and every time. Further, potassium hydroxide is corrosive and must be respected when worked with. In contrast, plasma etching (see Figure 1.13b) exposes the glass to intense oxygen plasma to remove organic debris. It uses atmospheric nitrogen as its gas source and, therefore, produces no chemical waste. In addition to producing less waste, plasma cleaning can produce a higher number of cleaned slides and

cover slips in a shorter amount of time compared to chemical etching. A typical protocol for plasma cleaning is listed below.

1. Scrub glass slides with a test tube brush in soapy water to remove any remaining epoxy or double sided tape from previous use.

2. After rinsing, sonicate slides and cover slips for 10 minutes in deionized (DI) water.

3. Remove slides and cover slips from the water and place them into the plasma cleaner.

4. Seal the door and power on the vacuum (but not the radio frequency source) for 10 minutes to dry the slides and cover slips.

5. Power on the radio frequency source for 10 minutes.

After cleaning, flow cells can be passivated and functionalized before or after flow cell construction depending on the desired method. Proper surface functionalization is important because it ensures the molecule of interest is immobilized in a controlled manner while passivation protects against non-specific interactions with the surface. There are three surface passivation and functionalization methods used in this thesis (see Figure 1.14):

1. Biotinylated bovine serum albumin (BSA). After cleaning, flow cells are constructed. Before an experiment, $50\,\mu\mathrm{L}$ of $0.2\,\mathrm{mg/mL}$ biotinylated BSA is flown into the cell and incubated for 5 minutes.

2. PEGylation. $80\,\mathrm{mg}$ of mPEG-silane and $2\,\mathrm{mg}$ of biotin-PEG-silane are suspended in $100\,\mathrm{mM}$ sodium bicarbonate solution. $75\,\mu\mathrm{L}$ are pipetted onto a clean slide. A cover slip is carefully placed over the wetted slide so that no air bubbles are formed. After 1 hour incubation in the dark, slides and cover slips are separated and rinsed with DI water and dried with compressed air. Finally, flow cells are constructed.

3. Slides and cover slips are rinsed twice in hexane. $50\,\mu\mathrm{L}$ of dimethyldichlorosilane (DDS) is added to a slide and cover slip container that is completely filled with hexane to avoid air and moisture contact. The container is sealed and shaken for $1.5\,\mathrm{hr}$.

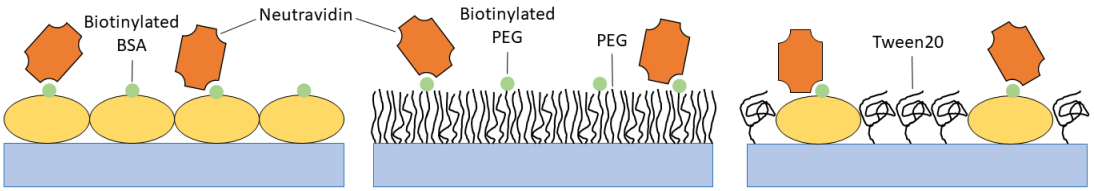


Figure 1.14: Surface passivation and functionalization of a cover slip. Left, biotinylated BSA. Middle, PEG. Right, Tween20.

Slides and cover slips are rinsed with fresh hexane and sonicated for $2\,\mathrm{min}$ in repetition for a total of 3 times. Slides and cover slips are air dried, and flow cells are constructed. Before the experiment, $50\,\mu\mathrm{L}$ of $0.2\,\mathrm{mg/mL}$ biotinylated BSA is flown into the cell and incubated for 5 minutes. After washing with T50 buffer (20 mM Tris and 50 mM NaCl at pH 8.0), $100\,\mu\mathrm{L}$ of 0.2% Tween20 in T50 is incubated for 10 minutes.

Later, all flow cells were rinsed with $50\,\mu\mathrm{L}$ of T50 followed by $25\,\mu\mathrm{L}$ of $0.2\,\mathrm{mg/mL}$ NeutrAvidin, which strongly binds to biotin at several sites and, therefore, can allow specific attachment of biotinylated DNA to the surface. Simple coating with BSA-biotin is the easiest and cheapest of the three methods; however, non-specific binding is the largest concern especially with cover slips and slides that have not been thoroughly cleaned. PEGylation and Tween20 offer consistent protection against non-specific binding; however, PEGylation is more costly. Quality of surfaces can tested for cleanliness by examination of the surface under laser illumination. Non-specific interactions can be quantified by counting spots after introducing $200\,\mathrm{nM}$ of a short fluorescently labeled oligo. Acceptable surface passivation should present as few spots as possible and certainly fewer than $\sim$25 spots.

Quantifying surface experiments

Experiments were digitally captured and recorded into binary files from an EMCCD camera with in-house Win32 Visual C++ software (see Figure 1.15). The program makes use of the camera manufacturer API as well as serial commands provided in the laser and pump operation manual to communicate with peripheral devices. The software includes helpful features for the experimenter such as file name protection which can change the current data file name to avoid overwriting previously written data with the same name. Further, the user can freely choose the working directory and save files in any location on the computer. Also, the software relays commands to both laser beams and the pump over an RS-232 serial data connection. The software allows for full control of the $532\,\mathrm{nm}$ and $640\,\mathrm{nm}$ lasers, including power/intensity regulation and the internal shutter. Finally, the syringe pump timing can be controlled which is important for experiments that require precise and accurate knowledge of the start time such as stopped-flow experiments.

The raw image data stored in the binary files present challenges when it comes to processing the data. For instance, an essential task when processing single-molecule experiments is to locate and identify spots in an image. This is a well-known problem in image processing, and there are even open source plugins available for the image processing program ImageJ [75] which are specifically dedicated to single-molecule data analysis [76]. Data in this thesis were processed with in-house MATLAB code. To prepare the image data, local background levels are calculated and subtracted from the image. We locate spots through a peak detection algorithm that counts peaks as individual threshold crossings that are well-separated from other intensity variations.

For FRET experiments, it is important to colocalize the image so that signals from different locations on the EMCCD (due to spectral separation) can be recognized as occurring at the same location in real space. This is accomplished by manually identifying the location of several control molecules where it is obvious that the signal in both spectral channels is from the same molecule. After selecting several sets of matching pairs, we can generate

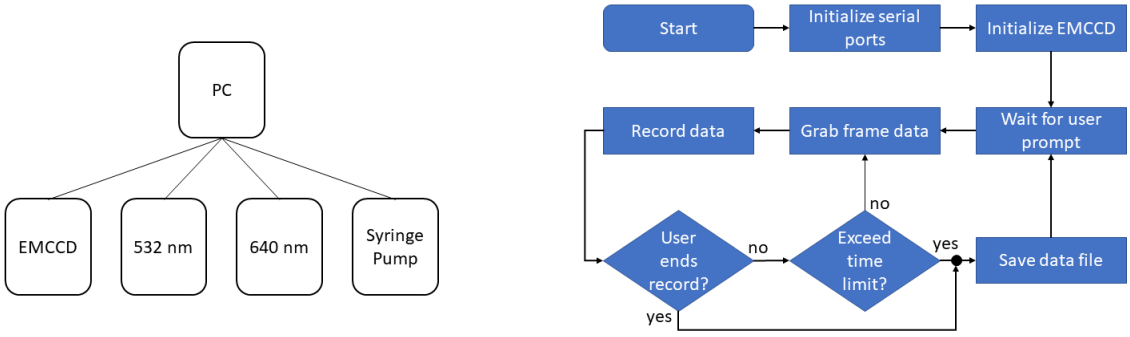


Figure 1.15: Instrumentation and camera software flowchart. On the left, the EMCCD camera, both laser beams, and the syringe pump are controlled by a single piece of software running on the computer. This allows for fast workflow and precise control over experiments especially ones that require precise timing of the connected instruments. On the right, the flowchart for the control software describes the data acquisition process. At start-up serial ports are initialized for communication with the lasers and pump. The EMCCD is initialized, and the software waits for a user prompt to grab and record data until either the user prompts to end acquisition or the predetermined time has been exceeded. This control loop can be modified to control the laser and pump behavior to suit the experimenter's needs.

an affine transformation, which does not preserve area but does preserve collinearity, between the image halves. This is accomplished with the fitgeotrans function in MATLAB. After candidate spots have been identified, time traces of fluorescent intensity for the spot location can be generated by calculating the sum of the surrounding pixel values for each frame of data. Traces are manually sifted to extract pertinent temporal information such as acceptor lifetime.

CHAPTER 2

ERASE: A NOVEL SURFACE RECONDITIONING STRATEGY FOR SINGLE-MOLECULE EXPERIMENTS

The last chapter provided an introduction to toehold-mediated strand displacement and described some tools that allow individual biomolecules to be located and observed. This chapter seeks to combine these two topics by demonstrating strand displacement in action at the single-molecule level. Toehold-mediated strand displacement reactions have been previously studied with single-molecule tools such as FRET and atomic force microscopy; however, the focus of these studies was not strand displacement itself. For example, some were to better understand DNA motors powered by strand displacement [77] or to observe large scale structural changes resulting from a strand displacement reaction [7].

This chapter's focus will be directed toward harnessing the power of strand displacement in single-molecule experiments. We will use single-molecule imaging to observe toehold-mediated strand displacement. In a reciprocal symbiosis, we will demonstrate how the strand displacement reaction can be used to enhance current methodologies in single-molecule fluorescence imaging by saving time spent for experiment preparation as well as decreasing the cost per experiment. Further, we will present an interesting application that will extend the throughput of surface-base single-molecule experiments.

As presented, this chapter is based on a manuscript in preparation for submission as:

D. B. Broadwater, R. B. Altman, S. C. Blanchard, and H. D. Kim, "ERASE: a novel surface reconditioning strategy for single-molecule experiments".

2.1 Abstract

While surface-based single-molecule experiments have revolutionized our understanding of biology and biomolecules, the workflow in preparing for such experiments, especially

surface cleaning and functionalization, remains labor-intensive and time-consuming. Even worse, meticulously assembled flow channels can be used only once for most experiments. A reusable surface would thus dramatically increase productivity and efficiency of single-molecule experiments. In this paper, we report a novel surface reconditioning strategy termed ERASE (Epitaxial Removal Aided by Strand Exchange) that allows a single flow cell to be used for vast repetition of single-molecule experiments. In this method, biomolecules immobilized to the surface through a nucleic acid duplex are liberated when a competing DNA strand disrupts the duplex via toehold-mediated strand displacement. We demonstrate the wide-range applicability of this method with various common surface preparation techniques, fluorescent dyes, and biomolecules including the bacterial ribosome. Beyond time and cost savings, we also show ERASE can assort molecules based on a nucleic acid barcode sequence, thus allowing experiments on different molecules in parallel. Our method increases the utility of prepared surfaces and is a significant improvement to the current single-use paradigm.

2.2 Introduction

Surface-based, single-molecule experiments comprise an essential toolkit to study biophysical mechanisms. In these experiments, the molecules of interest are tethered to a glass surface via a strong cohesive interaction between the surface and the molecule. Complexes that contain a nucleic acid component are often tethered via hybridization of the nucleic acid component to a single-stranded DNA bound to the surface. Such a tethering scheme has been used for many important biological systems, which include nucleosomes [78, 79], polymerases [80, 81], endonucleases [82], ribosomes [83, 84, 85, 86, 87], ribozymes [88], riboswitches [89], the DNA mismatch repair system [90, 91, 92], and CRISPR/Cas9 [93, 42].

The power of single-molecule investigation is best harnessed when molecules are housed inside a flow chamber rigged with an inlet and an outlet for liquid delivery. One of the ad-

vantages of this setup is the capability to change the solution phase by simple perfusion. Reversible reaction rates can then be measured from surface-bound molecules in various buffer conditions using the same flow channel. However, irreversible reaction steps such as high-affinity ligand binding and ATP or GTP hydrolysis can only be measured once from the molecules within the field of view, which constitutes only a tiny fraction of the flow chamber surface. In this regard, the flow chamber is heavily underutilized.

Similarly, host complexes of a different composition or sequence must often be investigated in separate flow channels [42, 94] unless they are modified with a different fluorescent label. Therefore, researchers have to prepare multiple flow channels to experiment with variants of the host complex or to obtain statistics for rate measurements. In our experience, preparing flow channels remains one of the most labor-intensive steps of the experimental procedure. A simple wash protocol that allows repeated usage of a single flow channel for different molecules would thus be highly beneficial to the single-molecule biophysics community.

In pursuit of such strategy, we noted that even a stable DNA duplex can be completely disrupted by a reaction called toehold-mediated strand displacement [94]. In this reaction, a single-stranded DNA molecule called the "invader" strand hybridizes to the substrate DNA strand in the toehold region and unzips the adjacent incumbent DNA strand by way of branch migration. We reasoned that this switch functionality could be harnessed as a surface reformatting tool, which would reduce the amount of time and materials necessary for preparing new flow cells.

Here, we introduce a novel surface-clearing strategy termed ERASE (Epitaxial Reformatting Aided by Strand Exchange). This method (see Figure 2.1(a)) involves three nucleic acid oligomers referred to as anchor, tether, and switch. The anchor is a short (10-15 nt) biotinylated oligomer that serves as the surface attachment site. The tether is the extension of the nucleic acid component of the complex of interest, which is purposefully designed to contain a terminal domain complementary to the anchor followed by a short (7-15 nt)

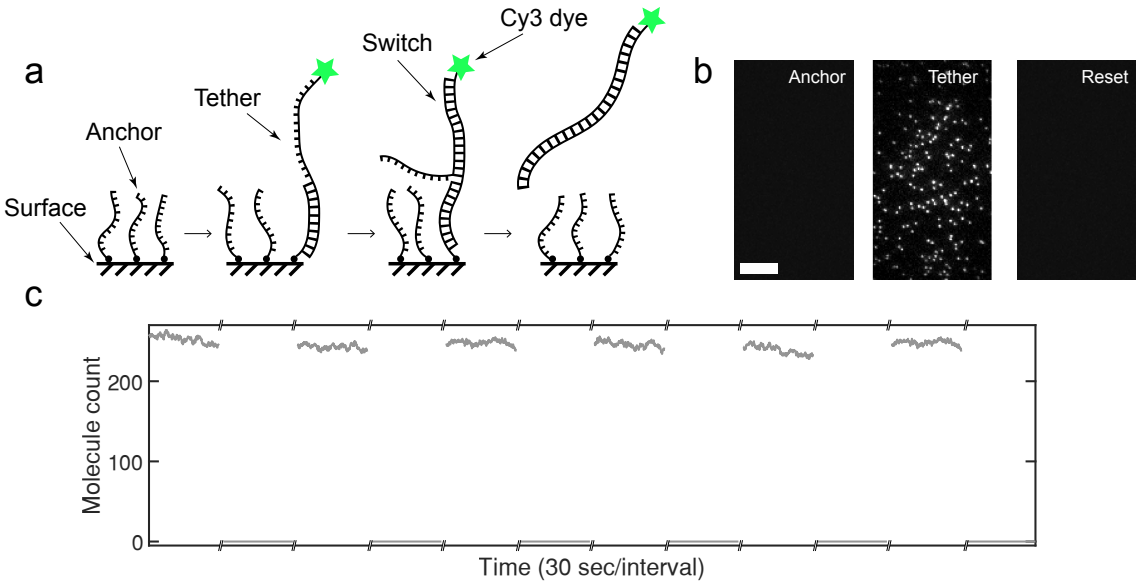


Figure 2.1: ERASE in action. (a) Biotinylated anchor molecules are immobilized on the surface. Cy3 labeled molecules with a tether are pumped into the flow cell and hybridize with the anchors. After reaching steady-state, switch molecules are flowed in. Strand displacement begins by binding of the switch to the open toehold region of the tether molecule. After branch migration, the anchor is displaced and the tether-switch duplex returns to solution. (b) After annealing to the anchor molecules, the tether molecules reach a steady-state surface density. The switch molecules entirely remove the anchor-bound tether molecules allowing the experiment to be repeated. The scale bar is $10\,\mu\mathrm{m}$. (c) Molecule counts over 30 second intervals from the same field of view are plotted. The high molecule count is maintained across several rounds of ERASE while the switch molecule efficiently removes all tether molecules.

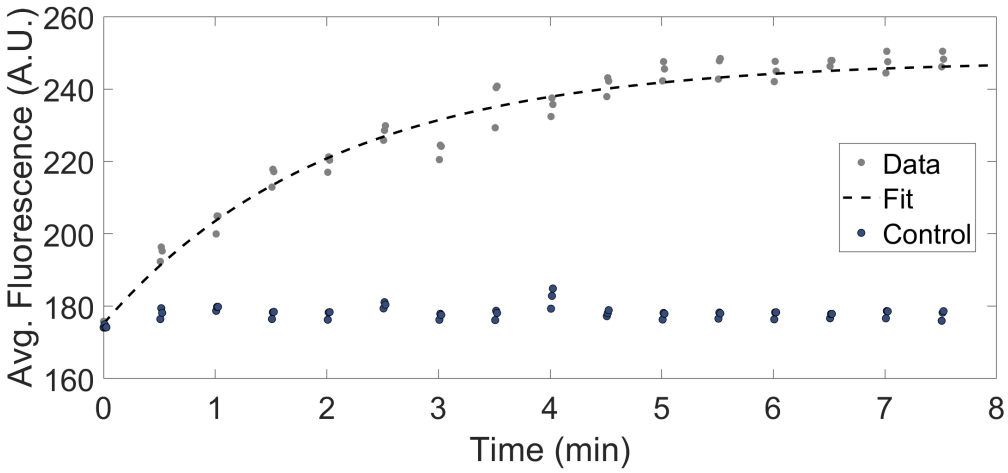


Figure 2.2: Binding rate to anchor molecules. We measured the mean fluorescence from Cy3 labeled DNA binding to anchor molecules on a PEGylated surface. To avoid photobleaching the laser is strobed for 2 second intervals every 30 seconds at an exposure time of 500 ms. The data is fitted with a single exponential to obtain a rate of anchor hybridization. As a control, Cy3 labeled DNA is introduced to a surface without anchor molecules.

spacer domain. The switch is fully complementary to the spacer and anchor domain of the tether molecule. According to this scheme, the experimental complex is immobilized through base pairing between the anchor and tether, while the addition of the switch removes the surface-tethered complexes through toehold-mediated strand displacement. After displacement, the surface is ready for another experiment by introducing molecules with the tether.

2.3 Methods

2.3.1 Sample Preparation

Custom nucleic acid oligomers were ordered from Integrated DNA Technologies and modified to include a Cy3 or Cy5 fluorophore near the $5'$ end or a biotin linker at either the $5'$ or $3'$ end. Unmodified switch molecules were ordered from Eurofins Scientific. The specific sequences are in Table 2.1. 70S ribosomal complexes were prepared as described previously [84].

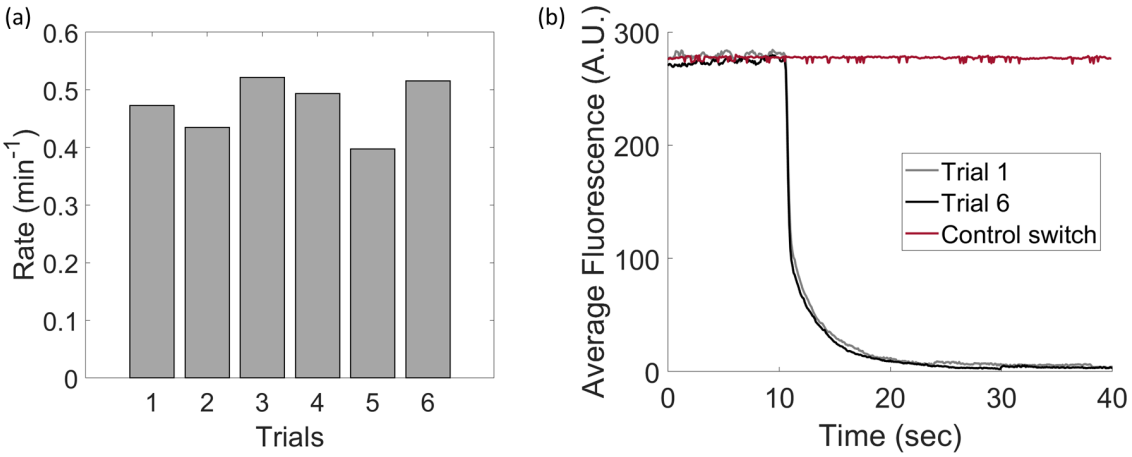


Figure 2.3: Consecutive rounds of ERASE. (a) We measured the rate of mean fluorescence over rounds of ERASE. Similar rate constants were extracted from single exponential fits of anchor hybridization for several consecutive rounds of ERASE. (b) We measured the rate of mean fluorescence for Cy3 labeled DNA after introducing the switch. The switch molecule is introduced at 10 seconds. Decrease in fluorescence is complete in tens of seconds and is similar from Trial 1 to Trial 6. A control switch with a different sequence does not engender a change in fluorescence.

Table 2.1: Oligonucleotide sequences used in this study.

Anchor 1	/BioTEG/CACTGGTGTT
Tether 1	CA/iCy3/ATTAAAATTCCGACAACACCAGGT
Switch 1	ACCTGGTGTTGTCGGAATTTTAAT
Control Switch 1	GAGTGGGGAGTCAAAGTAAATTCAAAACCAGGACTCAC TGCGAGGTA
Anchor 2	TCAATTCGTCGTC/BioTEG/
Tether 2	/Cy3/GACGACGAATTGAAGTGAAA
Switch 2	TTTCACTTCAATTCGTCGTC
RNA Tether 2	/Cy5/GACGACGAAUUGAAGUGAAA
DNA Anchor-70S	AGTTTTAGGTTGCCCCCCTTTTTTTTTTTTTTTTTTTTTT TTTTTT/BioTEG/
mRNA-70S	CAACCUAAAACUUACACACCCUUAGAGGGACAAUCGAU GUUCAAAGUCUUCAAAGUCAUC
Switch-70S	GGGTGTGTAAGTTTTAGGTTG
Orthogonal Tether	/Cy5/GACGACGAATTGATCACTTT
Orthogonal Switch	AAAGTGATCAATTCGTCGTC

2.3.2 Slide and cover slip cleaning and functionalization

8 holes were drilled by computer numerical control (CNC) across both long edges of a quartz slide using a diamond drill bit. Quartz slides and cover slips were sonicated for $10\,\mathrm{min}$ in Milli-Q water and dried by vacuum for $10\,\mathrm{min}$. Slides and cover slips were then placed in an upright position and plasma cleaned for $10\,\mathrm{min}$ (Harrick Plasma, PDC-32G). Flow cells were constructed by laying thin strips of double-stick tape across the slide. After aligning and pressing the cover slip against the slide, the open edges were sealed with 5-minute epoxy. Three surface functionalization methods were performed: 1) BSA, 2) PEGylation, and 3) DDS–Tween-20+BSA. 1) After slide construction, $25\,\mu\mathrm{L}$ biotinylated BSA ($1\,\mathrm{mg\,mL^{-1}}$) was incubated for $5\,\mathrm{min}$ in the flow cell. 2) Prior to flow cell construction, $2\,\mathrm{mg}$ of biotin-PEG-silane was mixed with $80\,\mathrm{mg}$ of mPEG-silane into $320\,\mu\mathrm{L}$ of $100\,\mathrm{mM}$ sodium bicarbonate solution. $80\,\mu\mathrm{L}$ of PEG solution was poured onto the quartz slide. The cover slip was placed on top of the slide and both were incubated at room temperature for an hour. The pair was disassembled, washed with Milli-Q water, and dried with compressed air. Flow cells were then assembled as described above. 3) After construction, slides and cover slips were rinsed in hexane and incubated in an upright container with a solution containing $75\,\mathrm{mL}$ hexane and $50\,\mu\mathrm{L}$ dichlorodimethylsilane for $1.5\,\mathrm{h}$ while gently shaking. The slides and cover slips were sonicated 3 times in fresh hexane for $2\,\mathrm{min}$ each. After slide construction, $25\,\mu\mathrm{L}$ biotinylated BSA ($1\,\mathrm{mg\,mL^{-1}}$) was incubated for $5\,\mathrm{min}$ in the flow cell. Then $100\,\mu\mathrm{L}$ 0.2% Tween-20 in T50 ($10\,\mathrm{mM}$ Tris-HCl, $50\,\mathrm{mM}$ NaCl, pH 8.0) was incubated for $10\,\mathrm{min}$ in the flow cell. After all 3 passivation techniques, the surface was neutravidin coated.

2.3.3 Experimental Setup

Objective-type total internal reflection fluorescence microscopy was performed to image single molecules with a commercial microscope (Olympus IX81). Cy3 labeled molecules were excited by $532\,\mathrm{nm}$ laser (NT66-968, B&W Tek), and Cy5 labeled molecules were

excited by $640\,\mathrm{nm}$ laser (CUBE 640-30FP, Coherent). Images were captured by EM-CCD (DU-897ECS0-#BV, Andor Technology) at $100\,\mathrm{ms}$ exposure and 2x2 binned by our in-house software. A syringe pump (NE-1000, New Era Pump System) was used to control flow volume and flow rate ($15\,\mu\mathrm{L\,s^{-1}}$). Nucleic acid molecules were pumped into flow cells in oxygen-scavenging imaging buffer containing $1\,\mathrm{mM}$ 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox), $5\,\mathrm{mM}$ protocatechuic acid, $100\,\mathrm{nM}$ protocatechuate 3,4-dioxygenase, $100\,\mathrm{mM}$ Tris-HCl (pH 7), and $1\,\mathrm{M}$ NaCl. Experiments with 70S ribosome complexes were performed in Tris-polymix buffer containing $50\,\mathrm{mM}$ Tris-acetate (pH 7.5), $5\,\mathrm{mM}$ $MgCl_2$, $100\,\mathrm{mM}$ KCl, $5\,\mathrm{mM}$ $NH_4(CH_3COO)$, $0.5\,\mathrm{mM}$ $CaCl_2$, $0.1\,\mathrm{mM}$ EDTA, $5\,\mathrm{mM}$ putrescine, and $1\,\mathrm{mM}$ spermidine in the presence of an oxygen-scavenging buffer consisting of $2\,\mathrm{mM}$ protocatechuic acid, $50\,\mathrm{nM}$ protocatechuate 3,4-dioxygenase, and $1\,\mathrm{mM}$ Trolox. The fluorescent signal was recorded and analyzed using in-house MATLAB software. Molecules were counted by peak detection by threshold.

2.4 Results

To demonstrate this method, we first tried a Cy3 labeled single-stranded DNA oligonucleotide containing a tether and a shorter biotinylated DNA strand as the anchor (see Figure 2.1(a)). The anchor molecules were introduced at $\sim 1\,\mathrm{nM}$, which would correspond to a density of $1\ \mathrm{anchor}/(120\,\mathrm{nm})^2$ at full adsorption. This surface density is high enough to provide ample binding sites for the tether, but low enough to prevent any nearest-neighbor interaction. The tether was then introduced at $50\,\mathrm{pM}$, which resulted in an increasing number of spots in the field of view. The time-dependence of this increase could be fit to a negative single exponential model (see Figure 2.2), which suggests that tethers bind to the surface via densely populated anchors in a pseudo first order process.

We next performed ERASE on the flow channel by introducing the switch at $500\,\mathrm{pM}$. The number of spots quickly decayed to the background level over a few seconds (see Figure 2.1(b), Figure 2.3(b)). As shown in Figure 2.1(c), the number of spots in the same field

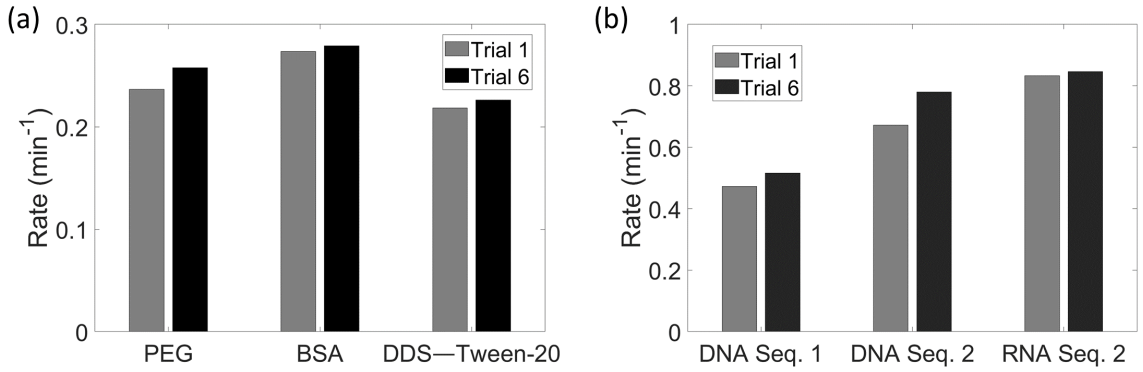


Figure 2.4: Binding rates for various surface and nucleic acids. (a) We measured the rate of mean fluorescence for Cy3 labeled DNA across different surface passivation schemes. Similar rate constants were extracted from single exponential fits of anchor hybridization for the first and sixth round of ERASE for three surface passivation techniques: 1) PEGylation, 2) BSA, 3) DDS–Tween-20+BSA. (b) We measured the rate of mean fluorescence for different sequences of Cy3 labeled DNA and Cy5 labeled RNA. Similar rate constants were extracted from single exponential fits of anchor hybridization for the first and sixth round of ERASE for different nucleic acids: 1) DNA with internally labeled Cy3, 2) DNA with a different sequence and end labeled with Cy3, 3) RNA with the same sequence as 2) and end labeled with Cy5.

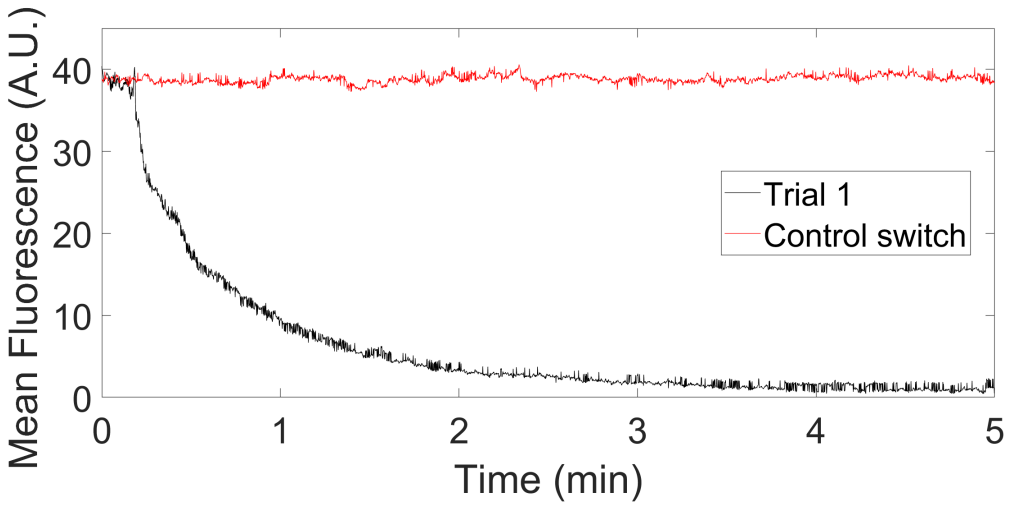


Figure 2.5: ERASE and 70S complex. We measured the rate of mean fluorescence for 70S complexes after introducing the switch. The switch molecule is introduced at 10 seconds. Decrease in fluorescence is complete in a few minutes. A control switch with a different sequence does not cause a change in fluorescence.

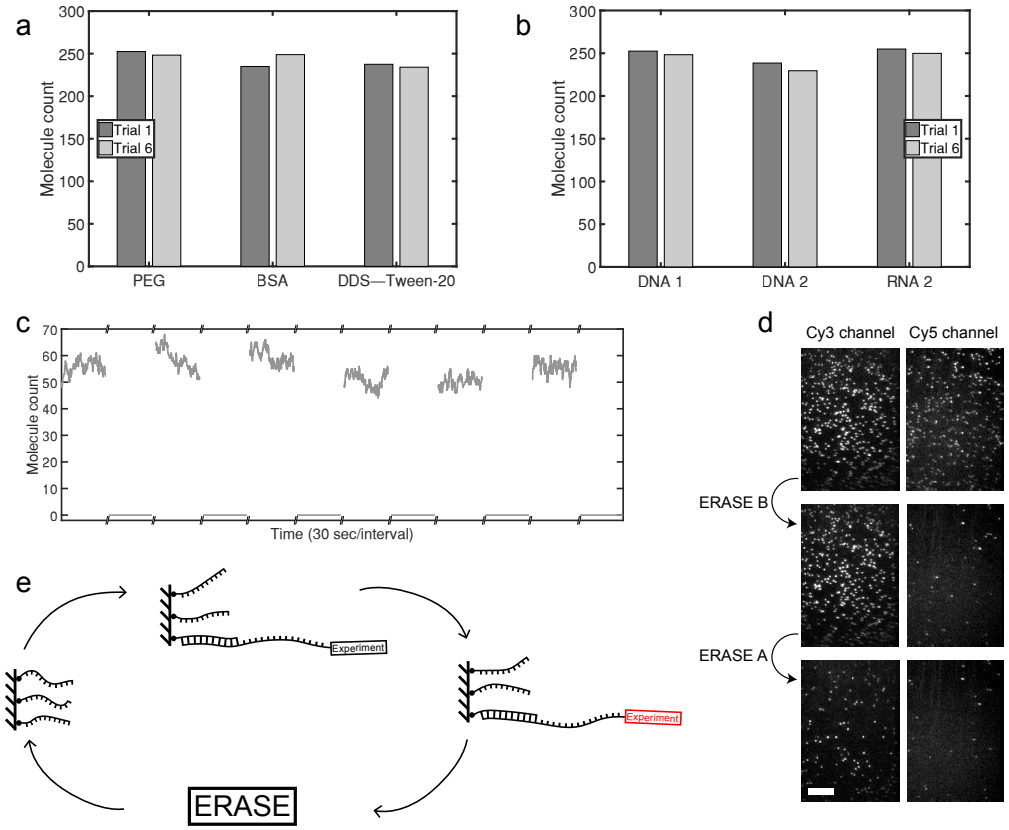


Figure 2.6: ERASE is versatile. (a) Molecule counts are repeatable and consistent across 3 different surface passivation techniques: PEGylation, BSA, and DDS–Tween-20+BSA. (b) ERASE is highly modifiable and can be constructed with any sequence demanded by the experiment. The molecule counts of a different anchor-tether-switch scheme perform comparably to the first scheme. ERASE is adaptable and can be augmented to different types of nucleic acid. Although it is composed of RNA, the tether molecules can still be entirely removed by the same DNA switch molecules. (c) Molecule counts of 70S ribosome complexes with an RNA tether are plotted, and ERASE efficiently removes all molecules. Molecule counts over the same field of view are consistent across several rounds of ERASE. (d) We designed two different DNA tethers, A and B: A is labeled with Cy3 dye and barcode sequence a, and B with Cy5 dye and barcode sequence b. Both tethers contain a common sequence complementary to the anchor. The Cy3- and Cy5-channel images of the same field of view are shown on the left and right columns, respectively. The first row of images are obtained after incubating the anchor-coated surface with a 1:1 mixture of A and B, the second row after ERASE with switch $\bar{\text{A}}$, which is complementary to A, and the third row after ERASE with switch $\bar{\text{B}}$, which is complementary to B. To minimize signal loss from the corresponding channel, we introduced switch molecules at a lower concentration which decreased ERASE efficiency. The scale bar is $10\,\mu\text{m}$. (e) ERASE is generalizable to any experiment using a nucleic acid tether. The tether-containing molecule can undergo irreversible changes (red) such as protein modification or annealing; however, provided the toehold remains intact, the switch molecule can always remove the tether-containing molecule.

of view changed from ~ 250 to nearly zero, indicating that ERASE completely removed all bound tether molecules. After ERASE, re-introduction of tethers at $50\,\mathrm{pM}$ could regenerate a similar density of spots in the same field of view. We performed this cycle of ERASE and tether addition as many as 6 times, and for each cycle, were able to achieve complete removal and recovery of spots in the same field of view with no sign of fatigue. The rates associated with removal and recovery also showed little difference between all trials (see Figure 2.3(a)). Finally, we found ERASE to be highly sequence-specific as the surface was reformatted only when the complementary sequence was used (see Figure 2.3(b)).

To showcase the wide adaptability of ERASE, we repeated the experiment with three common surface passivation techniques: PEGylation [95], bovine serum albumin (BSA) [94], and DDS–Tween-20+BSA [96]. We observed repeatable performance between trials that was consistent across all passivation schemes (see Figure 2.6(a), Figure 2.4(a)). ERASE is also adaptable to modifications in sequence as well as anchor and toehold length. We repeated the hybridization experiment with a separately designed scheme containing a different anchor length, toehold length, and sequence. We found the new Cy3 labeled DNA tether to produce a similar steady-state molecule count (see Figure 2.6(b)) as well as similar binding kinetics (see Figure 2.4(b)). Again, the switch molecule completely removed tethers from the surface. Finally, ERASE can be expanded to other nucleic acid types. We performed the experiment with a Cy5-labeled RNA tether with the same sequence as the second DNA tether except T-to-U substitution and observed comparable performance (see Figure 2.6(b), Figure 2.4(b)). Despite being a different nucleic acid, the same DNA switch molecule from the second design scheme efficiently removed the RNA tether.

The true power of ERASE lies in the ability to reprogram the surface for any biomolecules that harbor an end-exposed nucleic acid tether. To demonstrate this, we repeatedly ERASEd 70S ribosome particles (Cy3-Met-tRNAfMet in the P site) containing an mRNA tether. ERASE consistently removed all ribosomal complexes while molecule counts for each round were very similar (see Figure 2.6(c)). The switch molecule, designed to hybridize

the mRNA just beyond the anti Shine-Dalgarno sequence (see Table 2.1), specifically removed ribosomal complexes whereas a bioorthogonal control switch did not (see Figure 2.5).

Finally, we explored one possible application of ERASE for multiplexing. In this scheme, multiple types of molecules carrying different barcodes in the toehold region would be immobilized to the same flow channel and subjected to the same perfusion experiment. Subsequently, the positions of each molecule type could be determined by performing ERASE with barcode-specific switches in a sequential manner, allowing the mixed signal to be decomposed. As a proof of principle, we immobilized two different tether molecules fluorescently labeled with Cy3 and Cy5, respectively, and performed ERASE with one switch at a time. Each switch removed $\sim 95\%$ of its target molecules while non-target molecules were removed at a much lower rate of 1% (see Figure 2.6(d)). Optimizing the ERASE condition to maximize the false negative rate and minimize the false positive rate will be the topic of a future study.

2.5 Discussion

Apart from the time and cost savings due to repeated use of flow cells, ERASE provides several practical advantages to the experimenter. ERASE allows the same field of view to be reused across all experiments in case inhomogeneities from channel to channel or from field to field become an issue. Further, since switch molecules specifically disrupt base pairing between the anchor and the tether, spots that are not removed by switch molecules can be identified as nonspecifically bound or spurious and, hence, can be left out of data processing and statistics (see Figure 2.6(d)). Therefore, ERASE provides an a posteriori mechanism to select "good" molecules in an unbiased manner.

There are still unresolved questions surrounding this technique. Despite using a bioorthogonal switch molecule, a noticeable amount of cross-talk still exists between tether molecules. A possible extension of this method is to mutate the nucleotide in the tether of one molec-

ular species at the location most proximal to the toehold. The overall thermodynamic stability between the anchor and mutated tether would only be minimally impacted, but the kinetic discrimination of an invading switch molecule could be increased as this has been previously demonstrated in matching toeholds [46, 94]. However, given that the two tether species had entirely dissimilar toeholds, our finding seems to suggest an unknown method of strand displacement.

In conclusion, we introduced ERASE, a novel surface reformatting scheme that enhances the power of surface-based, single-molecule experiments. We have shown ERASE to be highly specific for removing the target molecules from the surface without compromising the long-term binding capacity of the surface for subsequent experiments. We have also shown it to be consistent and repeatable across common surface passivation techniques, and different sequences and types of nucleic acid. Due to its simplicity and flexibility, we expect ERASE to become routine for future single-molecule studies.

CHAPTER 3

THE EFFECT OF BASE PAIR MISMATCH ON DNA STRAND DISPLACEMENT

In Chapter 2, we demonstrated the feasibility of observing toehold-mediated strand displacement with single-molecule fluorescence techniques. Further, we exploited the switching behavior of toehold-mediated strand displacement to reformat surfaces and "wipe away" experiments that are attached via a nucleic acid tether. While ERASE was shown to be efficient and reliable, our exploration of the technique raised important questions with regards to the underlying mechanism of strand displacement. One among them asks what is the origin of the observed crosstalk between orthogonal switches and tethers? Such behavior is unexpected and cannot be accounted for with a current understanding of DNA strand displacement. Furthermore, in the field of DNA nanotechnology and engineering, strand displacement has a reputation as an ideal black box that outputs an incumbent strand upon input of a unique invader; however, as with all abstraction models, these black box models can fail spectacularly in circumstances that invalidate the underlying assumptions. Therefore, it is important to probe these assumptions and limits, so as to better improve the abstraction's realm of applicability.

Beyond the questions raised in the last chapter, an even more fundamental question surrounds the nature of strand displacement. Namely, how does a single base pair mismatch affect toehold-mediated strand displacement? To motivate this question, it is helpful to imagine how the thermodynamic stability of a duplex (either invader/substrate or incumbent/substrate) is impacted by a single base pair mismatch. A duplex of several base pairs is still quite stable in the presence of a mismatch. Clearly, toehold-mediated strand displacement should proceed even in the presence of a mismatch between the invader and substrate. It is unclear, however, what the effect is of a single base pair mismatch on the kinetics of toehold-mediated strand displacement. This chapter seeks to answer this question by using

single-molecule fluorescence imaging. We demonstrate the effect of secondary structure on strand displacement kinetics. Further, we develop two models to describe our observed behavior which point to a previously unconsidered mechanism of strand displacement.

Appendix A uses the methods developed in this chapter to examine more mismatch possibilities in an effort to explore alternative and exotic mechanisms of strand displacement. There, we uncover surprising results that corroborate the main conclusions of this chapter.

This chapter is based on contents of a manuscript which is published in full as:

D. B. Broadwater and H. D. Kim, "The effect of base pair mismatch on DNA strand displacement," *Biophysical journal*, vol. 110, no. 7, pp. 1476-1484, 2016.

As the author of the original article, I retain the rights to reproduce the article from the publisher. More information can be found at `https://www.elsevier.com/about/our-business/policies/copyright#Author-rights`.

3.1 Abstract

DNA strand displacement is a key reaction in DNA homologous recombination and DNA mismatch repair and is also heavily utilized in DNA-based computation and locomotion. Despite its ubiquity in science and engineering, sequence-dependent effects of displacement kinetics have not been extensively characterized. Here, we measured toehold-mediated strand displacement kinetics using single-molecule fluorescence in the presence of a single base pair mismatch. The apparent displacement rate varied significantly when the mismatch was introduced in the invading DNA strand. The rate generally decreased as the mismatch in the invader was encountered earlier in displacement. Our data indicate that a single base pair mismatch in the invader stalls branch migration, and displacement occurs via direct dissociation of the destabilized incumbent strand from the substrate strand. We combined both branch migration and direct dissociation into a model, which we term, the concurrent displacement model, and used the first passage time approach to quantitatively

explain the salient features of the observed relationship. We also introduce the concept of splitting probabilities to justify that the concurrent model can be simplified into a three-step sequential model in the presence of an invader mismatch. We expect our model to become a powerful tool to design DNA-based reaction schemes with broad functionality.

3.2 Introduction

DNA strand displacement is a reaction where one of the strands in a double-stranded DNA is replaced with another nearly identical strand. It is a fundamental mechanism to exchange genetic material and plays an essential role in homologous recombination [97] and mismatch repair [98, 99]. DNA strand displacement involves three single strands named the invader, the incumbent, and the substrate strands and can be abstracted to a swapping reaction between the invader and the incumbent strands on the substrate strand. The invader can then be viewed as an input signal, while the incumbent can be seen as an output signal. At this level of abstraction, DNA strand displacement can be idealized into "tinker toys" that fit together to form complex, interactive networks in the field of nanotechnology [100, 101, 102, 103] with applications in diverse areas such as biosensing [9, 10], DNA construction [5, 8, 6], DNA motors [104, 30, 105, 77, 106], and DNA computation [17, 21, 18, 16, 20, 19].

One class of strand displacement, known as toehold-mediated DNA strand displacement, is particularly useful because of sequence-dependent controllability. In this reaction, the shorter incumbent forms a partial duplex with the longer, complementary substrate (see Figure 3.1). The invader then hybridizes with the toehold, the unbound region of the partially-duplexed complement. The reaction is thought to proceed through a branch migration process until the incumbent is completely displaced [25]. The thermodynamics of this reaction is straightforward: the final state forms more canonical Watson-Crick base pairs and, therefore, must be lower in free energy than the initial state. In comparison, kinetics of strand displacement can vary by several orders of magnitude as a function of

toehold length [24] and mismatch position [46].

However, current models of DNA strand displacement are either too simplified [24] or too detailed [25, 46] to capture position-dependent sequence effects on strand displacement kinetics. This chapter seeks to build a reaction scheme for toehold-mediated DNA strand displacement kinetics at the single base pair level. To construct this model, we measured the strand displacement rate in the presence of a mismatched base pair in the invader and the incumbent using single-molecule fluorescence. We found that a mismatch in the invader could dramatically slow down the strand displacement rate when positioned near the toehold. Based on this observation, we devised a reaction scheme that includes both branch migration and direct dissociation of the incumbent, which can be analyzed with ease using the first passage time approach. The observed dependence of strand displacement rate on mismatch position suggests that a single mismatched nucleotide in the invader can stall branch migration, and direct dissociation of the incumbent, but not complete branch migration, terminates DNA strand displacement. Our model analysis thus reveals direct dissociation of the incumbent as an essential pathway of DNA strand displacement.

3.3 Materials and Methods

3.3.1 Sample preparation

Custom DNA oligomers were purchased from Integrated DNA Technologies (Coralville, IA), which were internally labelled near the $5'$ end with a Cy5 fluorophore to increase photostability [107] and a biotin linker at the $3'$ end for surface immobilization. The 26-nt sequence was chosen as a complement to a region of mRNA encoding Yellow Fluorescent Protein (YFP). The 14-nt incumbent sequences labeled with a BHQ-3 dark quencher at the 3' end were commercially synthesized by Biosearch Technologies (Petaluma, CA). The 24-nt invader sequences were purchased from Eurofins Scientific (Huntsville, AL). Single mismatch strands were chosen to preserve pyrimidine:purine ratio by exchanging $G \leftrightarrow A$ and $T \leftrightarrow C$. The specific sequences are in Tables 3.1, 3.2, and 3.3 shown below.

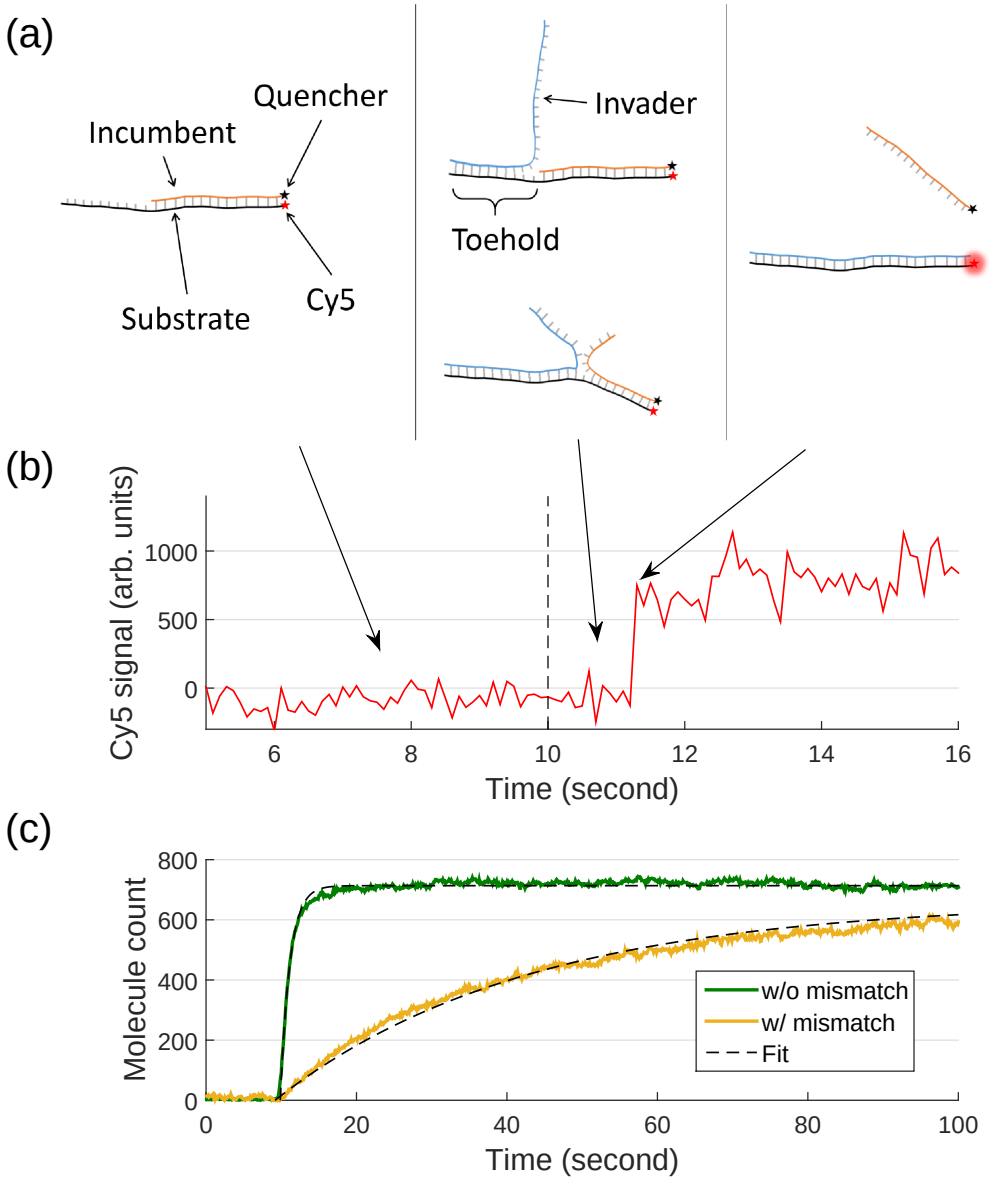


Figure 3.1: Measuring strand displacement. (a) Experimental design. Three single strands of DNA termed substrate (black), incumbent (orange), and invader (blue) strands participate in strand displacement. Cy5 attached to the substrate is initially quenched due to the Black Hole Quencher on the incumbent. When the incumbent is displaced by the invader, Cy5 recovers its fluorescence. (b) Cy5 signal during strand displacement. Shown is the fluorescence time trajectory of a single Cy5 molecule obtained by total internal reflection microscopy. Invader molecules were introduced via flow (dashed line at the 10th second). A large, single, and sudden increase in fluorescence indicates displacement. (c) Extracting the apparent strand displacement rate. Two sets of sample data and their respective fits are plotted. Molecule count is calculated via in-house code scripted in MATLAB (The MathWorks, Natick, MA). The data are fitted to single exponential curves with an origin at the injection time (10 seconds after starting acquisition). Mean first passage times (MFPTs) are approximated as the reciprocal of the fitted rate constants.

Table 3.1: Substrate strand sequences. The substrate strand sequence is complementary to a region of messenger RNA encoding YFP. We internally labelled the strand with Cy5 to increase photostability [107] and implemented a biotin linker for surface immobilization. The substrate* strand was derived from the substrate and altered to remove secondary structure.

Substrate	5'-CA/iCy5/ACCAAAATTGGGACAACACCAGTG/3BioTEG/-3'
Substrate*	5'-CA/iCy5/ATTAAAATTCCGACAACACCAGGT/3BioTEG/-3'

Table 3.2: Incumbent strand sequences. The incumbent strand was labelled with a dark quencher with an absorption spectrum that well overlaps the emission of Cy5. The underlined letter represents the single mismatch. The match* strand was designed to remove secondary structure.

Match	5'-GTCCCAATTTTGGT/BHQ3/-3'
Match*	5'-GTCGGAATTTTAAT/BHQ3/-3'
Mismatch 1	5'-A<u>A</u>TCCCAATTTTGGT/BHQ3/-3'
Mismatch 2	5'-G<u>C</u>CCCAATTTTGGT/BHQ3/-3'
Mismatch 3	5'-GT<u>T</u>CCAATTTTGGT/BHQ3/-3'
Mismatch 4	5'-GTC<u>T</u>CAATTTTGGT/BHQ3/-3'
Mismatch 5	5'-GTCC<u>T</u>AATTTTGGT/BHQ3/-3'
Mismatch 6	5'-GTCCC<u>G</u>ATTTTGGT/BHQ3/-3'
Mismatch 7	5'-GTCCCA<u>G</u>TTTTGGT/BHQ3/-3'
Mismatch 8	5'-GTCCCAA<u>C</u>TTTGGT/BHQ3/-3'
Mismatch 9	5'-GTCCCAAT<u>C</u>TTGGT/BHQ3/-3'
Mismatch 10	5'-GTCCCAATT<u>C</u>TGGT/BHQ3/-3'
Mismatch 11	5'-GTCCCAATTT<u>C</u>GGT/BHQ3/-3'
Mismatch 12	5'-GTCCCAATTTT<u>A</u>GT/BHQ3/-3'
Mismatch 13	5'-GTCCCAATTTTG<u>A</u>T/BHQ3/-3'
Mismatch 14	5'-GTCCCAATTTTGG<u>C</u>/BHQ3/-3'

3.3.2 Experimental setup

Objective-type total internal reflection fluorescence microscopy (TIRFM) was implemented to image individual molecules. A commercially available microscope (IX81, Olympus, Melville, NY) was used to image Cy5 fluorophores excited by a 640 nm laser (CUBE 640-30FP, Coherent, Santa Clara, CA). Binned images (2×2) were captured with an EMCCD (DU-897ECS0-#BV, Andor), and images were recorded at 10 fps with $100\,\mathrm{ms}$ exposure

Table 3.3: Invader strand sequences. The underlined letter represents the single mismatch. The strands marked with an asterisk were designed by removing secondary structure from their corresponding invader strands.

Match	5'-CACTGGTGTTGTCCCAATTTTGGT-3'
Match*	5'-ACCTGGTGTTGTCGGAATTTTAAT-3'
Mismatch 1	5'-CACTGGTGTTATCCCAATTTTGGT-3'
Mismatch 1*	5'-ACCTGGTGTTATCGGAATTTTAAT-3'
Mismatch 2	5'-CACTGGTGTTGCCCCAATTTTGGT-3'
Mismatch 3	5'-CACTGGTGTTGTTCCAATTTTGGT-3'
Mismatch 4	5'-CACTGGTGTTGTCTCAATTTTGGT-3'
Mismatch 5	5'-CACTGGTGTTGTCCTAATTTTGGT-3'
Mismatch 6	5'-CACTGGTGTTGTCCCGATTTTGGT-3'
Mismatch 7	5'-CACTGGTGTTGTCCCAGTTTTGGT-3'
Mismatch 7*	5'-ACCTGGTGTTGTCGGAGTTTTAAT-3'
Mismatch 8	5'-CACTGGTGTTGTCCCAACTTTGGT-3'
Mismatch 9	5'-CACTGGTGTTGTCCCAATCTTGGT-3'
Mismatch 10	5'-CACTGGTGTTGTCCCAATTCTGGT-3'
Mismatch 11	5'-CACTGGTGTTGTCCCAATTTCGGT-3'
Mismatch 12	5'-CACTGGTGTTGTCCCAATTTTAGT-3'
Mismatch 13	5'-CACTGGTGTTGTCCCAATTTTGAT-3'
Mismatch 14	5'-CACTGGTGTTGTCCCAATTTTGGC-3'

time using our in-house software. Experiments were performed on flow cells constructed as previously described [108], and a syringe pump (NE-1000, New Era Pump Systems) was used to control flow volume and flow rate ($10\,\mu\mathrm{L/s}$).

We performed experiments on flow cells that were created by placing double-sided tape between glass cover slips and glass slides to form several channels. Prior to assembly, we drilled holes into the glass slides followed by a cleaning protocol for both slides and cover slips. The cleaning protocol consisted of 20 minute sonication of slides and cover slips in $10\,\%$ Alconox, 20 minute sonication in purified water, 30 minute sonication in acetone, and a 30 minute sonication in 5 M KOH. After drying, the slides and cover slips were ready for assembly using double stick tape. We used 5-minute epoxy to seal openings and to affix 1/32 inch tubing to the drilled holes in the flow cells. We connected the tubing to a syringe pump (NE-1000, New Era Pump Systems) to control flow volume and flow rate

($600\,\mu\mathrm{L}/\,\mathrm{min}$).

The surface was passivated with biotinylated bovine serum albumin to minimize non-specific binding. After Neutravidin coating, the Cy5-labeled substrate molecules were immobilized at $50\,\mathrm{pM}$ in solution. Next, $20\,\mu\mathrm{L}$ of dark quencher-labeled incumbent strands were pumped into the flow cell at $200\,\mathrm{nM}$. After $5\,\mathrm{minutes}$, excess dark quencher probes were washed away with oxygen scavenging imaging buffer [109], which contained $1\,\mathrm{mM}$ 6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid (Trolox), $5\,\mathrm{mM}$ protocatechuic acid, $100\,\mathrm{nM}$ protocatechuate 3,4-dioxygenase, $100\,\mathrm{mM}$ Tris-HCl (pH 7), and $300\,\mathrm{mM}$ NaCl. Strand displacement was initiated by pumping invader strands in imaging buffer at $2\,\mu\mathrm{M}$ into the flow cell.

As the incumbent was displaced, fluorescent signal reappeared. The reappearance of fluorescent signal was recorded and analyzed using in-house MATLAB software. The molecule count (cumulative sum) was fitted to a single exponential curve and from that curve an overall rate of strand displacement was extracted. The mean first passage time was estimated as the reciprocal of the extracted rate. The experiment was repeated in triplicate for all single mismatch strands derived from the perfectly matched incumbent and invader.

3.3.3 Branch migration as random walk

We put forth a model for strand displacement based on the mean first passage time of a 1D random walk. We begin by assuming that the rate of breaking individual base pairs is much slower than the reverse rate of formation. By this assumption, incumbent strand unzipping and invader strand zipping is almost coincidental, and intermediates states can be specified with one state variable i, which is equal to the number of displaced base pairs. $i = 0$ is the initial state before invasion, and $i = n$ corresponds to complete displacement. We now define a 1D lattice with $n + 1$ sites. Motion at i-th lattice site is performed in single steps at forward and reverse rates, f_i and r_i respectively. This model is equivalent to a random walk with a perfectly reflecting boundary on the left ($i = 0$), and perfectly absorbing boundary

on the right ($i = n$).

The mean first passage time from $i = 0$ to $i = n$ is given by [110, 111]

$$\tau = \sum_{i=0}^{n-1} \frac{1}{p_i f_i}, \tag{3.1}$$

where p_i is the steady state probability at site i in a partial lattice between 0 and i. Therefore, the inverse of each term, $p_i f_i$, can be interpreted as the effective rate of reaching $i+1$ from an unspecified previous position. p_i can be expressed with a ratio of forward and reverse rates between two adjacent sites ($\alpha_i = f_{i-1}/r_i$) as

$$p_i = \frac{\alpha_i \alpha_{i-1} ... 1}{1 + \alpha_1 + \alpha_2 \alpha_1 + ... + \alpha_i \alpha_{i-1} ... 1} \tag{3.2}$$

Without sequence dependence, branch migration over a matched base pair must be identical in either direction and, therefore, $f_{i-1} = r_i$ or $\alpha_i = 1$. In comparison, $\alpha_i \gg 1$ for the case of a mismatch on the incumbent and $\alpha_i \ll 1$ for the case of a mismatch in the invader. We denote this mismatch-dependent fold-change in α as a, which must be larger than one for an incumbent mismatch and smaller than one for an invader mismatch. We also introduce variation in the forward rate for the first base pair to be displaced with another ratio (f/b). It is thought to be smaller due to slow initiation ($b > 1$) [112]. Using these ratios, the MFPT's with an invader mismatch (τ_v) and an incumbent mismatch (τ_c) at position j are given by

$$\tau_v(j) = \frac{1}{f}\left[-\left(\frac{1}{a} - 1\right) j^2 + n\left(\frac{1}{a} - 1\right) j + (b-1)\left(-\left(\frac{1}{a} - 1\right) j + \frac{n}{a}\right) + \frac{n(n+1)}{2}\right], \tag{3.3a}$$

$$\tau_c(j) = \frac{1}{f}\left[(1-a)j^2 - (1-a)(n+1)j + (b-1)((1-a)(j-1) + an) + \frac{n(n+1)}{2}\right], \tag{3.3b}$$

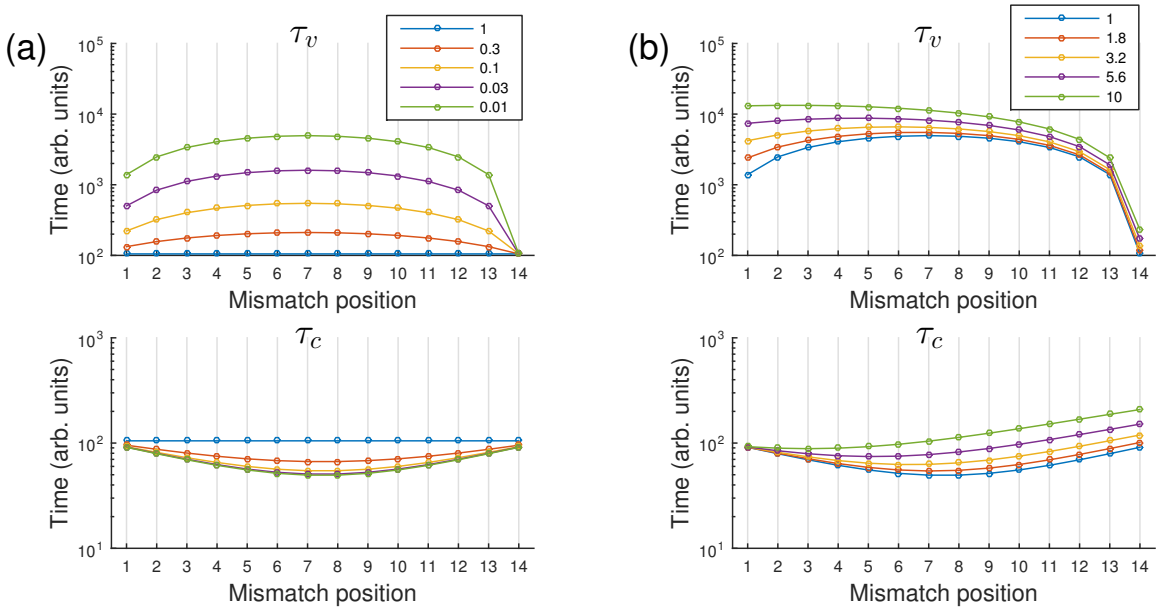


Figure 3.2: Dissociation via branch migration. MFPT's predicted by the 1D lattice model are plotted using Eq. 3.3. (a) MFPT's with varying a. a characterizes the effect of a mismatch on the forward branch migration rate. A mismatch in the invading strand lengthens MFPT (τ_v), while a mismatch in the incumbent strand shortens it (τ_v). b is fixed to 1. (b) MFPT's with varying b. b represents how slow the first migration step is compared to the rest. As b becomes larger, the center of the curves shifts towards the left, and both τ_v and τ_c become more monotonic as a function of mismatch position. a is fixed to 0.01.

respectively. The equations are cast in a form to reveal the dependence of MFPT on mismatch position j. Without the slower opening of the first base pair ($b = 1$), MFPT for the invader mismatch is concave down with a center at $n/2$, and MFPT for the incumbent mismatch is concave up with a center at $(n + 1)/2$. Slow opening of the first base pair ($b > 1$) shifts the center towards lower values. As expected, when $a = 1$ and $b = 1$, both MFPT's approach $n^2/2f$.

This 1D model predicts MFPT to be a quadratic function of mismatch position with the slowest displacement near the center position (see Figure 3.2(a)), which is not consistent with the overall monotonic change we observed with an invader mismatch. Slow initiation of branch migration ($b > 1$) could render the prediction more monotonic (see Figure 3.2(b)), but it requires an unreasonably large b.

In an effort to instill confidence in our model, we performed a separate experiment involving biotinylated invader strands, Cy3 labelled substrate strands, and Cy5 labelled incumbent strands. The invader strands were immobilized on the surface, and preformed substrate-incumbent duplexes were pumped in and allowed to react. Invader strands were designed to have the same complementary toehold (10nt) adjoined to a tail composed of 14 thymidines to prevent successful strand displacement. Interactions between duplexes and invader strands were recognized as a high FRET state. Lifetimes of high FRET states were recorded and interpreted as dissociation times. Large numbers of these lifetimes were recorded to construct a single exponential probability distribution whose decay rate was determined to be $\sim 0.03\,\mathrm{s}^{-1}$.

3.3.4 Concurrent displacement model

We define a 1D lattice with n sites, where n is the number of bases in the incumbent strand. For simplicity, we assume that the rate of breaking individual base pairs is slower than the reverse rate of formation. Under this assumption, the incumbent and the invader would remain completely zippered up with the substrate strand. Therefore, we can specify each intermediate state with one state variable i, which is equal to the number of displaced base pairs. For example, $i = 0$ represents the state where the invader has not displaced any base pair, and $i = n - 1$ corresponds to a state where the invader has displaced all but one base pair between the incumbent and the substrate. We add two boundary states (C and V) to this Markov chain. C stands for the in'c'umbent only state, and V for the in'v'ader only state. Branch migration at the i-th lattice site is performed in single steps at forward and reverse rates, f_i and r_i respectively. It is important to note that these rates are expected to be much slower than the single base pair opening (fraying) rate because a single fraying event does not necessarily lead to branch migration. As it stands, this model is equivalent to a random walk with a perfectly reflecting boundary on the left (C) and perfectly absorbing boundary on the right (V). Mean first passage time (MFPT) of this 1D model can be easily

derived with or without a kinetic barrier.

We can extend this model further to include direct dissociation of the invader and incumbent. As shown in Figure 3.3(b), the invader and the incumbent can dissociate at rates d_i' and d_i, respectively, from each state. This model scheme is thus similar to a general kinetic proofreading scheme [113] with additional feedforward paths to the final state. Given the toehold length of n_t, the invader is held by $n_t + i$ base pairs in state i and, therefore experiences a decrease in dissociation rate as more base pairs are formed. Conversely, since the incumbent is held by $n - i$ base pairs in state i, the dissociation rate would become larger as more base pairs are displaced. According to the previous work by Anshelevich *et al*. [114], the relationship between duplex dissociation rate ($k_d(N_{bp})$) and the number of base pairs (N_{bp}) is given by

$$k_d(N_{bp}) = \frac{2k_0 N_{bp}}{s^{N_{bp}-1}}, \tag{3.4}$$

where s is termed the stability factor equal to the ratio of rates of closing to opening for a single base pair, and k_0 is the unzipping rate of a single base pair at the melting temperature. This expression is essentially identical to the expression used by Zhang and Winfree [24].

The MFPT of the concurrent model requires consideration of the master equation

$$\frac{d\mathbf{x}}{dt} = -\mathbf{A}\mathbf{x}, \tag{3.5}$$

where $\mathbf{x}$ is defined as the $(n+1) \times 1$ state vector with probabilities in each state ($x_j(t)$) as components, and $\mathbf{A}$ is an $(n+1) \times (n+1)$ transition matrix, which is nearly tridiagonal, with components:

$$\mathbf{A} = \begin{bmatrix} k_a + d_0 & -d_0' & -d_1' & -d_2' & \cdots & -d_{n-2}' & -d_{n-1}' \\ -k_a & f_0 + d_0 + d_0' & -r_1 & 0 & \cdots & 0 & 0 \\ 0 & -f_0 & f_1 + r_1 + d_1 + d_1' & -r_2 & \cdots & 0 & 0 \\ 0 & 0 & -f_1 & f_2 + r_2 + d_2 + d_2' & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & & \cdots & f_{n-2} + r_{n-2} + d_{n-2} + d_{n-2}' & -r_{n-1} \\ 0 & 0 & 0 & 0 & \cdots & -f_{n-2} & r_{n-1} + d_{n-1} + d_{n-1}' \end{bmatrix}. \tag{3.6}$$

If the initial condition is given by $x_j(0) = \delta_{j1}$, the MFPT (τ) can be expressed with matrix determinants as [110]

$$\tau = \frac{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}}{\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}}, \tag{3.7}$$

where a_{ij} is the element of $\mathbf{A}$.

Without a mismatch, branch migration rate is the same in both directions ($f_{i-1} = r_i$). In comparison, a mismatch in the incumbent speeds up forward migration by a ratio a, and a mismatch in the invader speeds up reverse migration by the same ratio. a, termed the mismatch migration ratio should be much larger than one. Assuming that branch migration rates are identical to f for all nucleotides without a mismatch, the MFPT can be uniquely determined with five parameters k_a, k_0, s, f, and a.

3.3.5 Splitting probabilities

For the concurrent model, it is of particular interest to ask from which state the incumbent dissociates most frequently. This concept is related to splitting probabilities in stochastic processes. In our model, the incumbent strand can reach the absorbing boundary state (V) from $n + 1$ different states. The time dependence of V is given by the rate equation

$$\frac{dV(t)}{dt} = \sum_{j=1}^{n+1} k_j x_j(t) \tag{3.8}$$

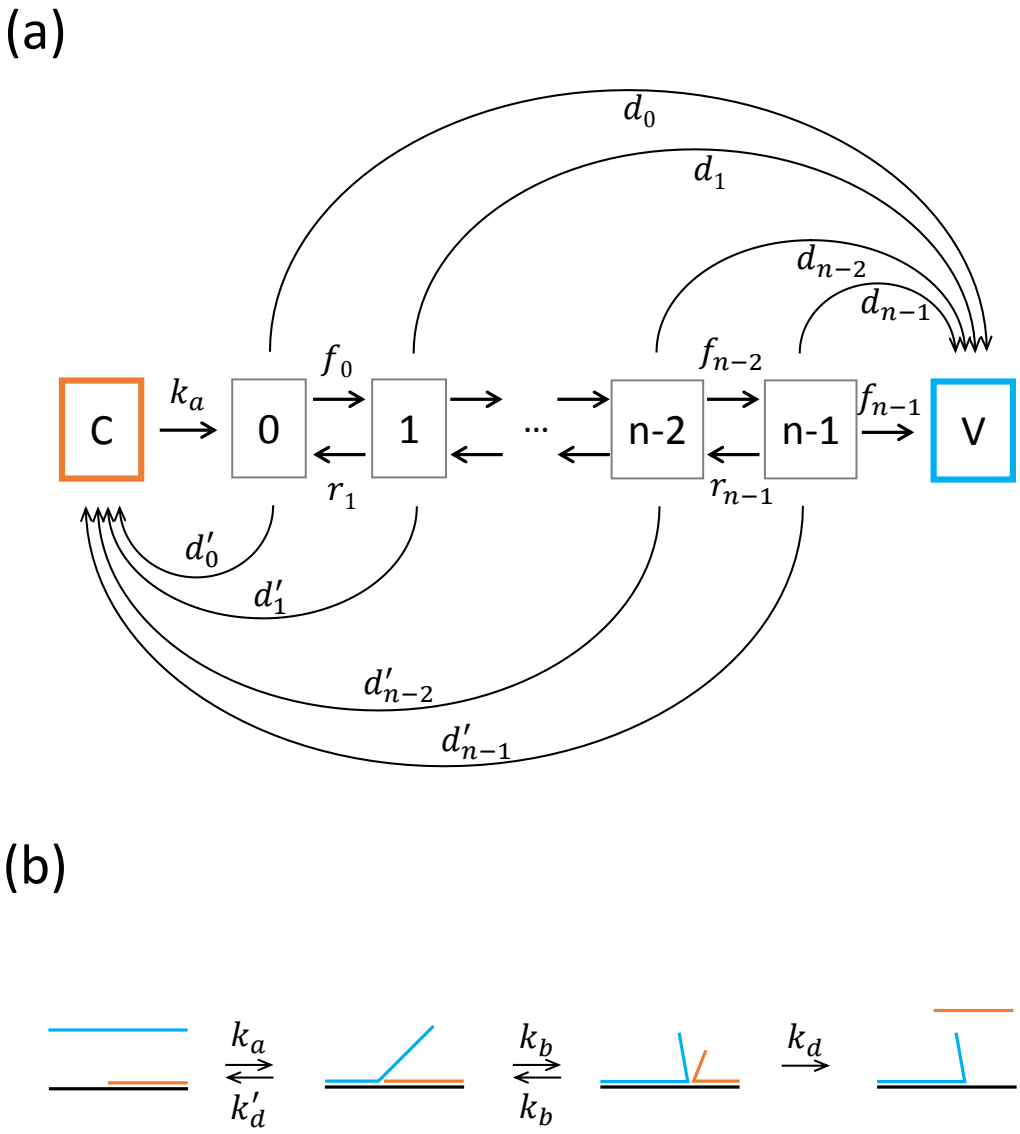


Figure 3.3: Strand displacement models. (a) Concurrent displacement model ("yin-yang" model). Strand displacement is preceded by a state occupied by the incumbent only (C). Branch migration begins after the toehold annealing step (k_a). The branch point can take any value (i) between 0 and $n-1$ and migrate towards nearest neighbors with forward rate (f_i) and reverse rate (r_i). V is the invader only state. Concurrently with branch migration, the invader and the incumbent can dissociate from any intermediate state with rates (d'_i and d_i) that depend on the number of base pairs . (b) Sequential displacement model. The initial state and the final state are identical to C and V. The orange, black, and blue line segments represent the incumbent, the substrate, and the invader, respectively. The invader may anneal (k_a) and dissociate (k'_d) with the toehold reversibly. Afterwards, branch migration (k_b) proceeds until a mismatch is encountered at state $j-1$ or the incumbent strand is significantly destabilized at state n_{th}. From either state, the incumbent can irreversibly dissociate (k_d).

where k_j is the dissociation rate constant from the j-th state. In the long time limit, the system will be completely depleted of the incumbent through $n+1$ channels.

$$V(t \to \infty) = \sum_{j=1}^{n+1} \int_0^\infty k_j x_j(t) dt \equiv \sum_{j=1}^{n+1} \pi_j, \tag{3.9}$$

where π_j is the splitting probability through the j-th state. Using eigen-decomposition of $\mathbf{A}$, it is straightforward to show

$$\pi_j = k_j \int_0^\infty x_j(t) dt = k_j \int_0^\infty \sum_{i=1}^{n+1} (e^{-\mathbf{A}t})_{ji} x_i(0) = k_j \sum_{i=1}^{n+1} (\mathbf{A}^{-1})_{ji} x_i(0), \tag{3.10}$$

where $\mathbf{A}^{-1}$ is the inverse matrix of $\mathbf{A}$. Thus, with the initial condition $x_j(0) = \delta_{j1}$, we obtain the following formula for splitting probabilities

$$\pi_j = k_j \frac{\mathbf{C}_{1j}}{|\mathbf{A}|}, \tag{3.11}$$

where $\mathbf{C}$ is the cofactor matrix of $\mathbf{A}$. π_j is also related to the mean first passage time (Eq. 3.7) according to

$$\tau = \sum_{j=1}^{n+1} \frac{\pi_j}{k_j}. \tag{3.12}$$

3.3.6 Sequential displacement model

One can also build a three-step sequential model (see Figure 3.3(b)) that qualitatively captures the effect of a mismatch on strand displacement. The first step is toehold formation through annealing (k_a) accompanied by reverse dissociation (k'_d). The second step is reversible branch migration (k_b). The third step is dissociation of the incumbent (k_d), which is irreversible in our experiment. The key difference of the model from the concurrent model is that branch migration and dissociation occur in a serial fashion. The mean first

passage time (τ) for this reaction is given by

$$\tau \approx \frac{1}{k_a} + \frac{1}{k_b} + \frac{2}{k_d}. \tag{3.13}$$

This equation can be derived from either Eq. 3.7 or Eq. 3.2 under the approximation that the invader association rate is faster than the dissociation rate ($k_a \gg k_d'$). The third step (incumbent dissociation) can occur from a state where (i) branch migration is stalled due to a mismatch in the invader, or (ii) the incumbent-substrate interaction is significantly weakened, with only a few intact base pairs left between them. We model that n_{th} number of base pairs have to be displaced for the incumbent to dissociate. The branch migration rate (k_b) also depends on migration distance (N_{bp}), which can be derived from the mean first passage time of a standard one-step process (Eq. 3.1) as

$$k_b(N_{bp}) = \frac{2f}{N_{bp}(N_{bp}+1)}. \tag{3.14}$$

If the mismatch in the j-th position is encountered before n_{th}, branch migration stalls at state $j-1$. In the absence of a mismatch, branch migration continues till the threshold state n_{th}. Whichever occurs first becomes the state where the incumbent dissociates ($\min(j-1, n_{th})$). Therefore, we can express the dependence of Eq. 3.13 on mismatch position as

$$\tau = \frac{1}{k_a} + \frac{1}{k_b(\min(j-1, n_{th}))} + \frac{2}{k_d(n - \min(j-1, n_{th}))}. \tag{3.15}$$

3.3.7 Data fitting

Nonlinear least squares fitting was performed with 'lsqcurvefit' of the MATLAB Optimization Toolbox. Eq. 3.7 was used as the fitting function. All individual measurements were fitted with equal weight using shared fitting parameters. These measurements include mismatch in the invader and the incumbent as well as the perfect match strand.

3.4 Results

We performed toehold-mediated strand displacement by challenging a surface-immobilized substrate-incumbent partial duplex with invader strands free in solution. In this experimental scheme, every reaction step can be treated as first order. Formation of partial duplexes between the substrate and the incumbent on the surface led to disappearance of most Cy5 spots due to quenching. Upon perfusion of the invader, Cy5 spots reappeared over time, which was interpreted as strand displacement. We counted individual spots over time and extracted the apparent displacement rate from single exponential fitting. We performed this experiment in triplicate for 15 invaders (14 mismatch strands + match strand). In Figure 3.4(a), we plotted the strand displacement rates measured for each mismatch position. It took $\sim 2\,\mathrm{s}$ for a perfectly matching invader to displace the incumbent (red point, Figure 3.4(a)). When the mismatch was introduced in the invader, strand displacement became slower, especially for the first four positions near the toehold region. The relationship was overall monotonic (except strands 7 through 10) with a roughly 70-fold change in the observed rate between the strands with a mismatch in the first and last positions. The effect of the invader mismatch is the strongest at the first position, but seems to be significantly weakened by position 6.

We suspected that the deviation from this trend at positions 7 through 10 might stem from a secondary structure in the invader. Strand 7 and 8, for example, are predicted to form stable hairpins (see Figure 3.5). Thus, we designed new sequences free of secondary structure for another set of strand displacement experiments. The invaders we tried were perfect match and mismatches at position 1 and 7. The rate measured with the new mismatch 7 strand was significantly faster, similar to the baseline, whereas the rates measured with the new perfect match and mismatch 1 strands remained unchanged. This result lends support to our speculation that the deviant points are caused by secondary structure.

We performed a similar experiment to explore the relationship between displacement

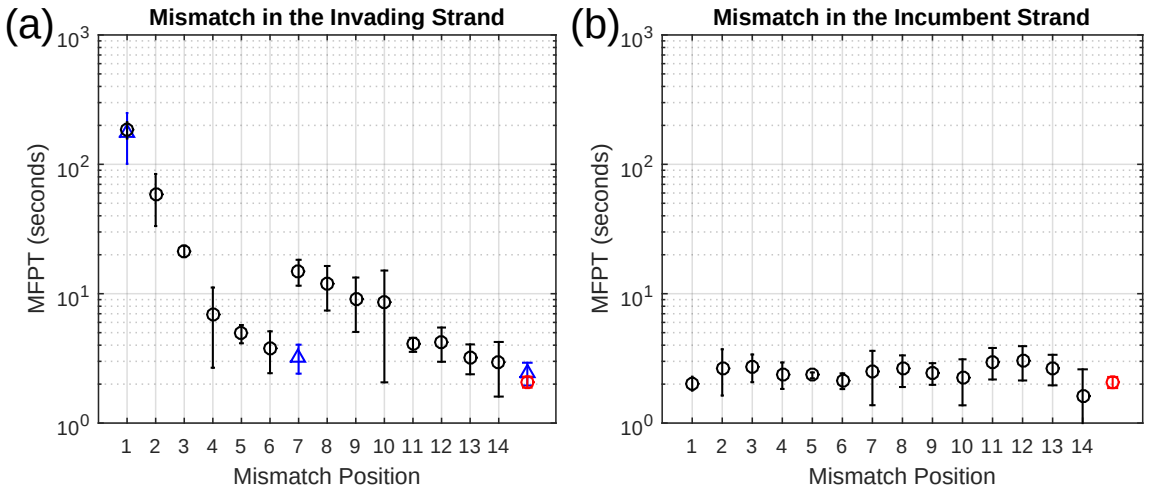


Figure 3.4: Observed displacement MFPTs. Molecule number was counted as a function of time and fitted to a single rate, and MFPT was calculated as the reciprocal of that rate. We plot the average and standard deviation of observed MFPT against the mismatch position in both the invader and incumbent. The perfect match case is plotted in red. The strands designed to be free of secondary structure are plotted in blue. (a) Invader mismatch. The MFPT generally increases with mismatch position and shows a nearly 70-fold variation overall. Notably, at positions 7, 8, and 9, the MFPTs are higher than the overall trend. The MFPTs for strands without secondary structure align closely with their counterparts for the match strand and mismatch position 1 strand, but there is a clear distinction for the mismatch position 7 strands. (b) Incumbent mismatch. In comparison to (a) the mismatch position does not have a significant effect on the MFPT.

rate and single mismatch position on the incumbent strand. In Figure 3.4(b), we plotted the measured strand displacement rates for each mismatch position. In contrast to the dynamic pattern for the invader mismatch, there is relatively little variation in displacement rate against the mismatch position in the incumbent strand. All rates were similar to the displacement rate without a mismatch (red point, Figure 3.4(b)). This incumbent mismatch experiment serves to control for the possibility of interacting dangling ends since the same dangling ends are available to interact in both the invader and incumbent mismatch. The lack of variation in rate over mismatch position for the incumbent implies that the observed complex behavior for the invader mismatch (see Figure 3.4(a)) is not due to interacting dangling ends.

These two sets of experiments corroborate the intuition that it is easier to displace a

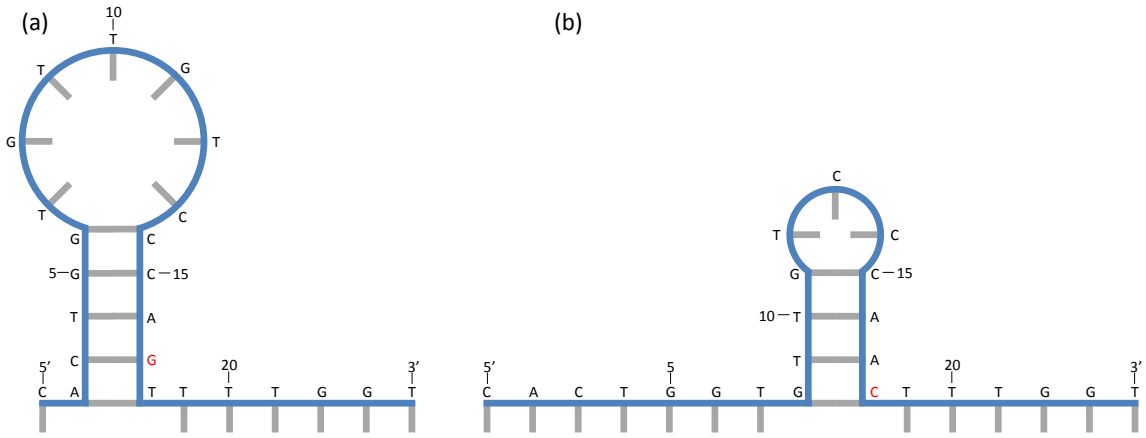


Figure 3.5: Putative secondary structures of invader mismatch strands. (a) Mismatch position 7. This is the only conformation predicted by mfold [115] for this structure. The hairpin mostly obstructs the toehold. It is predicted to have a lower free energy than the active form by $\Delta G = -3.61\,\mathrm{kcal/mol}$. (b) Mismatch position 8. This is the predicted conformation by mfold. The toehold is partially obstructed due to the hairpin. It is predicted to have a lower free energy from the active form by $\Delta G = -2.61\,\mathrm{kcal/mol}$.

mismatch strand with a match strand than *vice versa*. Also, it confirms the previous inference of branch migration rate of $1\,\mathrm{s}^{-1}$ over a similar length of displacement domain [24]. In contrast, a mismatch in the invader can dramatically slow down strand displacement, especially, if placed near the toehold. This implies that the mismatch effect could be modeled as a localized kinetic barrier that disrupts zipping of the invader.

To understand the mismatch effect in a quantitative fashion, we attempted to model toehold-mediated strand displacement as a one-dimensional random walk [111, 116] with a single misstep. This model assumes that displacement of the incumbent occurs via complete branch migration towards the boundary. The branch point moves much faster forward upon an incumbent mismatch and much faster backward with an invader mismatch. Derivation of the mean first passage times is straightforward for both invader and incumbent mismatches, and the analytical formula are presented as Eq. 3.3. However, this model incorrectly predicts a parabolic dependence of rate on mismatch position, with the mismatch in the center having the most significant effect [117]. While further inclusion of a slow initiation step causes the model to produce a general monotonic trend (see Figure 3.2),

it still cannot produce the sharp drop in rate observed over the first few positions.

The failure of the one-dimensional model prompted us to extend the model by including direct dissociation of the incumbent even before the branch point reaches the end. We reasoned that these direct dissociation paths would become important especially when branch migration is hindered or when the incumbent binding is severely weakened near the latter stage of branch migration. This model is termed the concurrent displacement model, and is schematized in an aesthetic yin-yang pattern as shown in Figure 3.3(a).

We performed nonlinear least squares fitting of MFPT (Eq. 3.7) to the measured displacement times. The dissociation rate (d'_0) of the invader from the toehold was directly measured (see Figure 3.6) and constrained in the fitting. All other invader dissociation rates ($d'_{i=1,2,\ldots,n-1}$) were expressed in relation to the measured value d'_0 by a single parameter s according to Eq. 3.4. The 4 data points at mismatch positions 7, 8, 9, and 10 that markedly deviated from the monotonic pattern were omitted, which we justify based on our additional measurement with strands rationally designed to be free of secondary structures (blue triangles, Figure 3.4). As shown in Figure 3.7(a), the concurrent displacement model can well fit both observed relationships with a common set of parameters. The mismatch migration ratio (a) diverges, and therefore an upper bound was placed. The best fit produces the association rate (k_a) of $0.6\,\mathrm{s}^{-1}$, the dissociation rate constant (k_0) of $3 \times 10^5\,\mathrm{s}^{-1}$, the branch migration rate (f) of $10\,\mathrm{s}^{-1}$, and the stability factor (s) of 5.1.

The association rate, $\sim 0.6\,\mathrm{s}^{-1}$ at $\sim 2\,\mu\mathrm{M}$, is similar to the association rate constant measured in bulk ($\sim 1\,/\mu\mathrm{M/s}$) considering the surface effect [118] or differences in salt condition or temperature [24, 119]. This value is also close to the association rate ($1\,\mathrm{s}^{-1}$ at $2\,\mu\mathrm{M}$) inferred from our separate measurement of concentration dependence (see Figure 3.6). The extracted branch migration step time is $100\,\mathrm{ms}$. This is seemingly much longer than $2.5\,\mathrm{ms}$ previously inferred based on the three-step displacement model [24]. This disparity, however, is not due to different measurements of apparent branch migration rates (k_b), but likely due to different models used to infer the step rate (f). For example, the

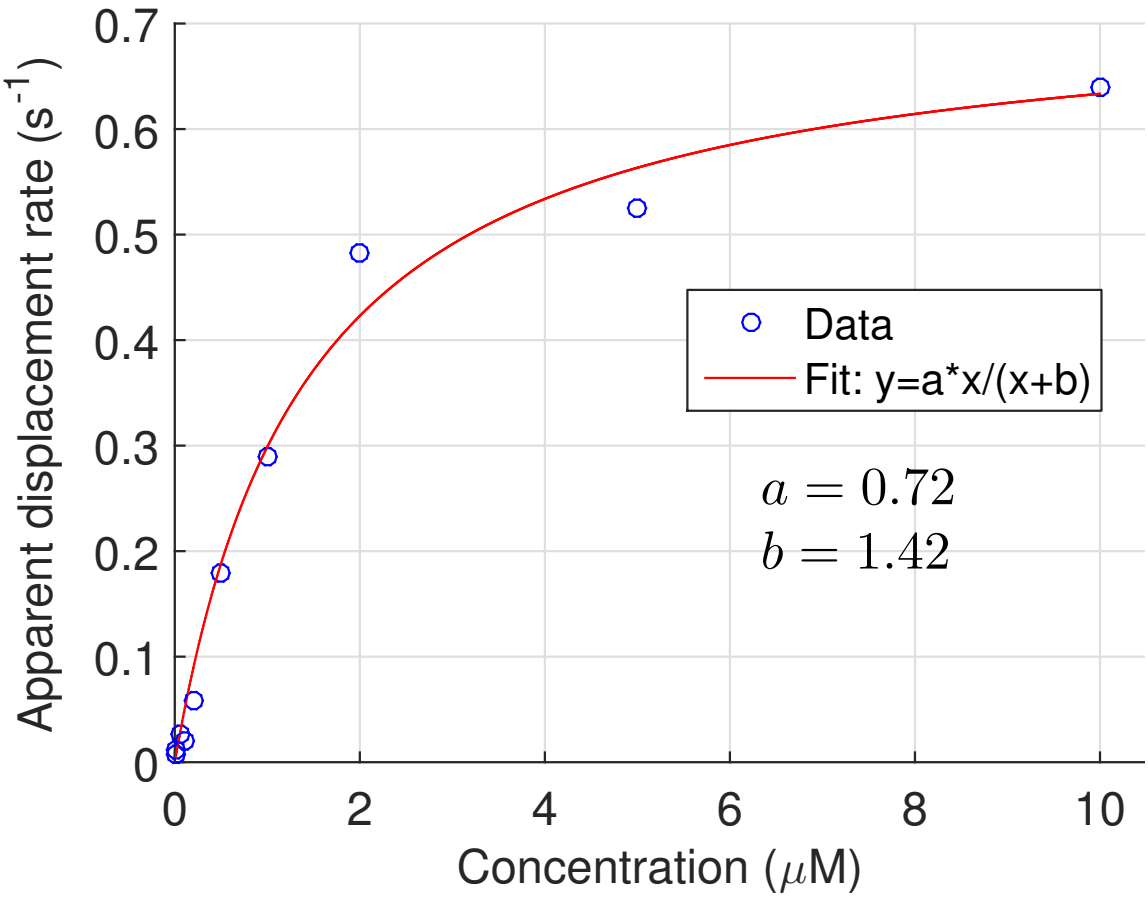


Figure 3.6: Displacement rate vs. invader concentration. The displacement rate (y) was measured as a function of the concentration of the invader strand (x). The displacement reaction can be modeled by reversible binding (k_+) and unbinding (k_-) steps followed by a unimolecular displacement (r) step. The unbinding rate of the toehold-bound invader was directly measured to be $1/30\,\mathrm{s}^{-1}$, which is much slower than r. In this case, the apparent displacement rate is given by $rx/(x + r/k_+)$. We fit the measured data points (blue hollow circles) using the expression $y = ax/(x + b)$ with two fitting parameters. From a and b, we determine the unimolecular displacement rate (r) and the binding rate (k_+) to be $0.72\,\mathrm{s}^{-1}$ and $0.5\,/\mu\mathrm{M}/\mathrm{s}$, respectively.

apparent time it takes to displace a 14-nt domain in our experiment is $\sim 1\,\mathrm{s}$ (red point in Figure 3.4)), similar to the inferred branch migration rate of $\sim 1\,\mathrm{s}$ over a 20-nt domain [24]. The spontaneous unzipping rate of a single base pair (k_0) is estimated to be $10^6\,\mathrm{s}^{-1}$ to $10^7\,\mathrm{s}^{-1}$ [114, 25]. Our estimate of $3 \times 10^5 s^{-1}$ is within an order of magnitude, and is also similar to a thermodynamic estimate ($6 \times 10^5\,\mathrm{s}^{-1}$) used by Zhang and Winfree [24]. Finally, the extracted stability factor (s) is 5.1, which indicates that the base pair is 5.1 times more likely to close than open. This ratio is close to $10^{0.6}$ obtained by extrapolation of a semi-analytical calculation [120].

To gain more insights into the mismatch effect, we calculated the probability that the incumbent strand dissociates from each state using the parameters obtained from fitting to the concurrent model. This probability is conceptually similar to the splitting probabilities between different absorbing states in a one step process [121]. In our concurrent model, the splitting probabilities leading to the absorbing state V can be calculated using Eq.3.11. In Figure 3.7(b), the splitting probability for each state i is plotted as a bar graph with varying mismatch positions marked by red vertical lines. As expected, the splitting probabilities sum to one in all cases. For early mismatch positions (left half, Figure 3.7(b)), splitting probabilities past the mismatch position are zero, which indicates that branch migration does not proceed beyond the mismatch. For late mismatch positions (right half, Figure 3.7(b)), the incumbent dissociation is complete even before the mismatch is encountered, which explains why the displacement rate is not affected by the mismatch. The key insight from this model analysis is that the invader mismatch stops branch migration, and displacement is terminated by incumbent dissociation, not by branch migration. Based on this insight, we can build a simpler sequential displacement model (see Figure 3.3(b)) to rationalize the observed dependence of strand displacement rate on mismatch position. The MFPT of this reaction scheme is expressed as a sum of three terms, association time, branch migration time, and dissociation time (Eq. 3.15). The position dependence mainly arises from the third term, which decreases with increasing mismatch position only up to

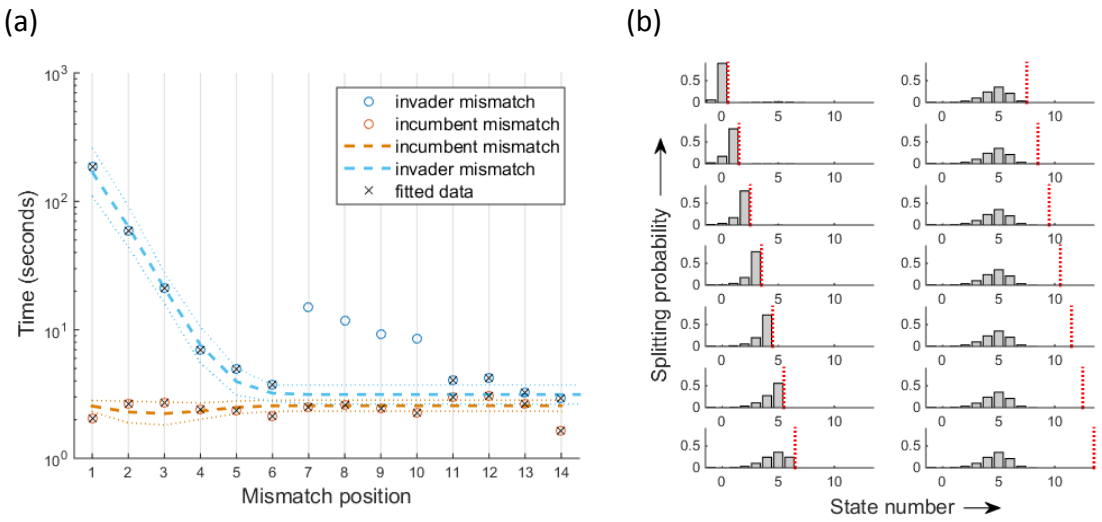


Figure 3.7: Model analysis. (a) Fit for concurrent displacement model. The plot shows the relationship between the apparent mean displacement time vs. the mismatch position in the invader (blue) and the incumbent (orange). We used five fitting parameters, the annealing rate (k_a), the dissociation rate constant (k_0), the branch migration rate (f), the stability factor (s), and the mismatch migration ratio (a). The points included in the fitting routine are marked by '$\times$'. The dashed lines show the fit by our model, and the dotted lines are $95\,\%$ confidence intervals at the input values. (b) Splitting probability distribution. The bar graphs show the splitting probability vs. state number with the mismatch position varying from 1 to 7 (top to bottom, left), and from 8 to 14 (top to bottom, right). The parameters obtained from the fit in (a) were used to calculate the probabilities. The red vertical dash indicates the mismatch position.

some threshold state (n_{th}) and remains unchanged beyond it.

3.5 Discussion

In this chapter, we designed a surfaced-based single-molecule assay to measure kinetics of toehold-mediated DNA strand displacement and its dependence on a base pair mismatch. In contrast to bulk measurements [24, 25, 46], our assay produces the strand displacement rate from a first-order reaction, which does not depend on substrate concentration. Furthermore, due to the long toehold and high invader concentration used in our assay, strand displacement is completed in a time scale of a few seconds (or minutes with a mismatch), significantly faster than typical bulk experiments. Due to this high efficiency, we expect our

experimental method to become a powerful tool for the study of the biophysics of branch migration. Although not exploited in this chapter, the single-molecule aspect of our method can produce the full distribution of individual strand displacement times as well [113, 122], which will be a topic of the next chapter.

Recently, the effect of an invader mismatch was studied in bulk [46]. This study found a similar qualitative dependence of displacement rate on mismatch position to ours, and hinted direct incumbent dissociation as an important pathway for displacement. The authors made an extra effort to preserve the trinucleotide sequence around a mismatch to minimize variation in mismatch free energy, which led to omission of some mismatch positions including the first. A qualitative explanation based on dynamics simulation was given, but a quantitative model predictive of displacement rates was missing. Our study thus complements the previous study by testing the mismatch effect in both the invader and the incumbent, at different positions, and with different DNA sequences. More importantly, we present a quantitative model and a first passage time approach to rationalize the mismatch effect.

Our concurrent model uses an intermediate level of coarse-graining compared to two types of previous models for toehold-mediated strand displacement. The first type is the three-step displacement model, which breaks the reaction into bimolecular toehold binding, unimolecular branch migration, and unimolecular dissociation from the final state [24]. The second type is a more detailed model at the molecular level, which includes intermediate states during branch migration [25]. Because the three-step model coarse-grains all of branch migration into a single step, it cannot readily incorporate nucleotide-level effects. The second model is thorough, but the implementation and analysis of this model require dynamics simulations with constraints and pre-equilibrium assumptions [46]. In comparison, our concurrent displacement model is straightforward to analyze using the first passage time approach presented here.

A complete understanding of the concurrent model requires solving the master equation

(Eq.3.5). However, because the system has many intermediate states with transition rates of similar magnitudes, the time dependence of the probability distribution ($x_j(t)$) is expected to be characterized by multiple exponential terms, and cannot be easily subjected to fitting analysis. The first passage time (MFPT) approach bypasses this technical difficulty. Unlike $x_j(t)$, calculation of the MFPT can be easily done without solving the master equation. Moreover, splitting probabilities can be easily obtained as well. The quantitative framework we employed here can thus be applied to any other complex reaction scheme.

In fitting a five parameter model to 24 data points, we recognize that precise values of each parameter become difficult to determine. This limited range of data points is inherent to the nature of the experiment. A larger data set requires a longer invader, which would become more susceptible to secondary structure formation and more spurious intermolecular interactions. Further, sequence dependence in individual steps of branch migration can lead to deviations from our model prediction. Nevertheless, fitting parameters k_0 and s are in agreement with other studies, and $k_a = 0.6\,\mathrm{s}^{-1}$ is similar to our own estimate of $1\,\mathrm{s}^{-1}$ (see Figure 3.6). Based on the extracted k_0 and s, the dissociation rate of the incumbent from state 4 is predicted to be $2.6\,\mathrm{s}^{-1}$. If complete base pairing of the substrate is assumed, state 4 corresponds to 10 base pairs between the incumbent and the substrate. But the dissociation rate of the 10-bp duplex between the invader and the substrate was measured to be much slower at $1/30\,\mathrm{s}^{-1}$. This comparison suggests that in state 4, the incumbent-substrate interaction is markedly destabilized probably due to a repulsive interaction between the incumbent and the invader near the branch point [123]. Interestingly, Srinivas *et al*— [25] found that branch migration intermediates are destabilized by $3.4k_BT$ due to dangling ends, which corresponds to $\sim$ 30-fold change in dissociation rate. Our model analysis is thus consistent with the incumbent having effectively 2-3 fewer intact base pairs than indicated by the location of the branch point. In other words, the incumbent is not completely zippered up against the invader.

Our own estimation of branch migration time of $\sim 100\,\mathrm{ms}$ per base pair step, notwith-

standing measurement and fitting uncertainty, is much slower than predictions in the literature based on base pair fraying and one dimensional rate models [24, 25]. But, we think our estimation is plausible for a few reasons. A single fraying event is not likely to provide enough room or time for a base on another strand to invade. Thus, invasion of a base *in trans* should occur only after many fraying/unfraying events, which could explain our slow branch migration rate. Also, 1D rate models are expected to significantly overestimate migration rates by omitting direct dissociation pathways which are the dominant mechanisms of displacement. To the best of our knowledge, branch migration step time has not been measured directly in the specific context for strand displacement. The branch migration time in Holliday junctions is estimated to be faster at $3.6\,\mathrm{ms}$ [124], but it is not accompanied by a pair of free dangling ends that can destabilize the branch point due to crowding [25] or thermal fluctuation [125]. Also, the low ionic strength used in our experiment may reduce branch migration rate as low salt would lead to slower base pair formation [119]. A different experimental strategy that prevents or decouples direct dissociation of the incumbent is certainly necessary to accurately measure the branch migration rate.

We used high concentrations of invader strand to speed up strand displacement and to minimize variability in our measurement. We measured the displacement rate for the perfect match invader as a function of concentration (see Figure 3.6) and selected a concentration significantly above the midpoint ($\sim 2\,\mu\mathrm{M}$). In this regime, the displacement rate is relatively insensitive to variation in concentration, and allows us to compare rates with different invaders. It was recently shown that DNA duplex can dissociate by competing complementary single strands without toehold mediation [126], but this effect kicks in at a much higher concentration ($\sim 50\,\mu\mathrm{M}$). Furthermore, the rates that we observed are at least an order of magnitude faster. Therefore, this mechanism cannot be relevant to our observations.

We have not comprehensively investigated the origin of deviation seen in invader mismatch strands 7,8,9, and 10. These strands exhibit significantly slower displacement rates

than the rest, which led us to consider secondary structure formation. Using the secondary structure prediction program mfold [115], we find that invader strands 2, 3, 7, and 8 adopt a relatively stable secondary structure (see Figure 3.5). Such secondary structure can negatively impact strand displacement at the annealing step. In addition to internal secondary structure, individual strands can form transient base pairs with one another, which could also retard the branch migration rate. The reduction in rate due to secondary structure would be more noticeable for 7 and 8 where strand displacement is fast. Furthermore, measurements for strands designed without secondary structure were much closer to the expectation (see Figure 3.4 blue). We could not account for the origin of outliers at position 9 and 10 (see Figure 3.4).

Our method is not without limitations. First, we infer MFPT by exponential extrapolation from the appearance of fluorescence signal. In theory, MFPT cannot be measured if the initial population size is not known or if the reaction is not complete. Nonetheless, all displacement kinetics curves and their exponential extrapolations have similar integrated areas (see Figure 3.1(c)), which indicates that our MFPT estimation is accurate. Second, Cy5 and the quencher in our experimental design might stack with each other or with neighboring bases to affect the intrinsic dissociation kinetics of the incumbent strand. However, such stabilizing interaction would only attenuate position-dependent mismatch effect, if any. Moreover, a similar experimental design did not affect the apparent displacement rate [46]. Therefore, the main conclusion we draw based on our model is likely valid.

Our results have interesting implications for related areas in biology. Given the role of strand displacement in homologous recombination, it is conceivable that the repair mechanism could be affected by a single base mismatch in a position dependent manner. In a more applied sense, position dependence of strand displacement rate could be exploited to design masked probes for single-nucleotide variant detection in vitro or in vivo with increased specificity.

3.6 Conclusion

We used a novel experimental strategy to study toehold-mediated DNA strand displacement as a first-order reaction. At the single-molecule level, we measured the apparent displacement rate through recovery of fluorescence and found its strong dependence on mismatch position in the invader. We rationalized the observed dependence using the concurrent displacement model that allows branch migration and dissociation of the incumbent. Our model analysis suggests that a single base pair mismatch in the invader poses an almost insurmountable kinetic barrier to branch migration and reveals direct dissociation of the destabilized incumbent as the dominant pathway for displacement. We anticipate our kinetic model, which we colloquially term the yin-yang model, and the first passage time approach to be highly relevant to an understanding of dynamic response for an expansive range of complex networks.

CHAPTER 4
PROBING THE KINETICS OF BRANCH MIGRATION

In the last chapter, we explored the kinetics of toehold-mediated DNA strand displacement. As shown in Figure 3.6, the measured displacement rate is highly dependent on concentration. This data highlights the kinetic dominance of a concentration-dependent, bimolecular step which is clearly the binding of the toehold. The dependence of strand displacement kinetics on toehold binding is of great interest to the field of DNA nanotechnology and great progress has been made in this regard [24, 25]. In comparison, relatively little is known about the kinetics of branch migration which is typically modeled as a 1D symmetric random walk; however, Chapter 3 highlighted the relative importance of branch migration and raised questions regarding the mechanism of branch migration.

In this chapter, we explore the kinetics of strand displacement by developing a single-molecule FRET technique to answer fundamental questions about branch migration. This experimental work involves several challenges that must be overcome. One of which is capturing the fluorescent signal from the extremely short timescale of branch migration on an EMCCD camera. In typical single-molecule imaging experiments, the EMCCD exposure time is around $100\,\mathrm{ms}$; however, in this chapter, we will explain how we are able to capture video images at an exposure time that is 25 times smaller. The results of this work point to a previously unknown factor in branch migration kinetics.

This chapter is based on a manuscript in preparation for submission as:

D. B. Broadwater and H. D. Kim, "Probing the kinetics of branch migration".

4.1 Abstract

Nucleic acid strand exchange processes are pervasive in biology and DNA nanotechnology. Toehold mediated DNA strand displacement is an archetypal strand exchange process

that proceeds through two distinct sequential mechanisms: toehold formation and two-way branch migration. As toehold formation is bimolecular and concentration-dependent, it dominates the kinetics in bulk which obfuscates the kinetic details of two-way branch migration which is unimolecular and fast for length scales of interest. We used single-molecule Fluorescence Resonance Energy Transfer to decouple these two bimolecular and unimolecular timescales which allowed us to directly observe the first passage time of branch migration. This assay allows us to measure important details, such as salt, temperature, and sequence, that, previously, have not been directly measured. We report significant sequence dependence for branch migration time. Further, we measured the branch migration time for the complementary invader of observed sequences. Despite the high degree of symmetry between such conjugate systems, we observed a significant difference in branch migration times. We constructed a model using the free energy of dangling ends that justifies the observed differences in relative invasion times between conjugate sides of invasion.

4.2 Introduction

Nucleic acids' ability to form hydrogen bonds between complementary Watson-Crick bases allows for a rich set of complicated, multi-step kinetic behaviors which include duplex hybridization [127], Holliday junction structural dynamics [128, 129], strand invasion [3], and two-way branch migration [94], of which the latter three are strand exchange processes. These strand exchange processes span a wide range of biological phenomena encompassing homologous recombination and CRISPR/Cas homology recognition. Further, DNA nanotechnology exploits this complex behavior to pioneer intricate device construction by creating nanoscale boxes with lock and key [4], cargo sorting robots [28], and logic gates [33].

Toehold-mediated DNA strand displacement is the prototypical example of such strand exchange processes as it includes only three single DNA strands and is not protein-mediated. It is initiated by bimolecular annealing between an invader strand and substrate strand at a

region known as the toehold. The reaction then proceeds through a unimolecular process known as branch migration which is the exchange of base pairs between the invader and incumbent along the substrate. Fundamentally, these reactions are thought to proceed in single steps of pairwise opening and closing reactions with a characteristic time of $\sim 10\,\mu$s [130, 131, 25].

Great attention has been paid to the kinetics of toehold-mediated strand displacement [94, 25, 19, 132]; however, current experimental efforts do not separate the bimolecular toehold formation kinetics from the branch migration kinetics. Bulk methods have estimated the branch migration rate on long ($\sim 1\,\mathrm{kbp}$) DNA [131], but these length scales extend far beyond the length scales of interest for DNA digital circuits [19] and CRISPR/Cas systems [42]. To our knowledge, there is no direct measurement of two-way branch migration kinetics.

In this chapter, we propose an approach based on single-molecule Fluorescence Resonance Energy Transfer (smFRET) to directly measure the first passage time of branch migration. In this approach, we immobilize the invader so that the toehold is proximal to the surface. We then introduce previously assembled partial duplexes which are composed of an acceptor-labeled (Cy5) incumbent annealed to a donor-labeled (Cy3) substrate (see Figure 4.1). After a diffusive search in bulk, the partial duplex rapidly anneals to the toehold and branch migration, indicated by a high FRET signal, proceeds. Upon completion of branch migration, the incumbent strand is fully displaced and a low FRET signal is observed.

With this strand displacement assay, we can observe distributions of branch migration passage time. We found that branch migration time varies significantly across different sequences. Furthermore, we found that branch migration times are not equal when invading from conjugate (mirror image) sides which suggests dangling ends to play a significant role in branch migration kinetics. Using previously observed dangling end free energy parameters, we constructed a rate model that correctly predicts the relative rates of invasion

Table 4.1: Invader sequences. The first 12 sequences correspond to the sequences in Figure 4.2(c) as read from top to bottom and then left to right. The loaded invaders and RNA sequences correspond to partial duplexes formed from the first two rows of Tables 4.2 and 4.3.

5′-/BioTEG/ACCTGGTGTTGTCGGAATTTTAAT-3′
5′-ATTAAAATTCCGACAACACCAGGT/BioTEG/-3′
5′-/BioTEG/ACCTGGTGTTTGTTCGTACGTGTG-3′
5′-CACACGTACGAACAAACACCAGGT/BioTEG/-3′
5′-/BioTEG/GGTTATTGGGTTGGGTGTGTGAAA-3′
5′-TTTCACACACCCAACCCAATAACC/BioTEG/-3′
5′-/BioTEG/CAATCAAATAAAACTCTCTCTAAA-3′
5′-TTTAGAGAGAGTTTTATTTGATTG/BioTEG/-3′
5′-/BioTEG/AGGTGGTGGTCTTTTAGGAGGAAA-3′
5′-TTTCCTCCTAAAAGACCACCACCT/BioTEG/-3′
5′-/BioTEG/ACCTGCTGTTGAACTATAGTCGTC-3′
5′-GACGACTATAGTTCAACAGCAGGT/BioTEG/-3′
5′-/BioTEG/ACCTGGTGTTGTCGGAATTTTAATTTTTT-3′
5′-/BioTEG/ACCTGGTGTTGTCGGAATTTTAATTTTTTTTTTT-3′
5′-/BioTEG/ACCUGGUGUUGUCGGAAUUUUAAU-3′
5′-AUUAAAAUUCCGACAACACCAGGU/BioTEG/-3′

between corresponding sides.

4.3 Materials and Methods

4.3.1 Sample Preparation

Custom DNA oligomers were purchased from Integrated DNA Technologies. The 26-nt substrate was internally labeled near the end distal to the toehold with a Cy3 fluorophore to increase photostability [107]. The 24-nt invader molecule was labeled with a BioTEG linker at the end proximal to the toehold for surface immobilization. The 14-nt incumbent sequences were labeled with a Cy5 fluorophore at the end distal to the toehold. All oligos were HPLC purified. The specific sequences are in Tables 4.1, 4.1, and 4.3. Partial duplexes were constructed by combining substrate and incumbent at a 1:1 ratio of $20\,\mu\mathrm{M}$ each in $1\,\mathrm{M}$ NaCl and $10\,\mathrm{mM}$ Tris at pH 7. The mixture was heated to $95\,^{\circ}\mathrm{C}$ and slowly cooled for $3\,\mathrm{h}$ to $4\,^{\circ}\mathrm{C}$.

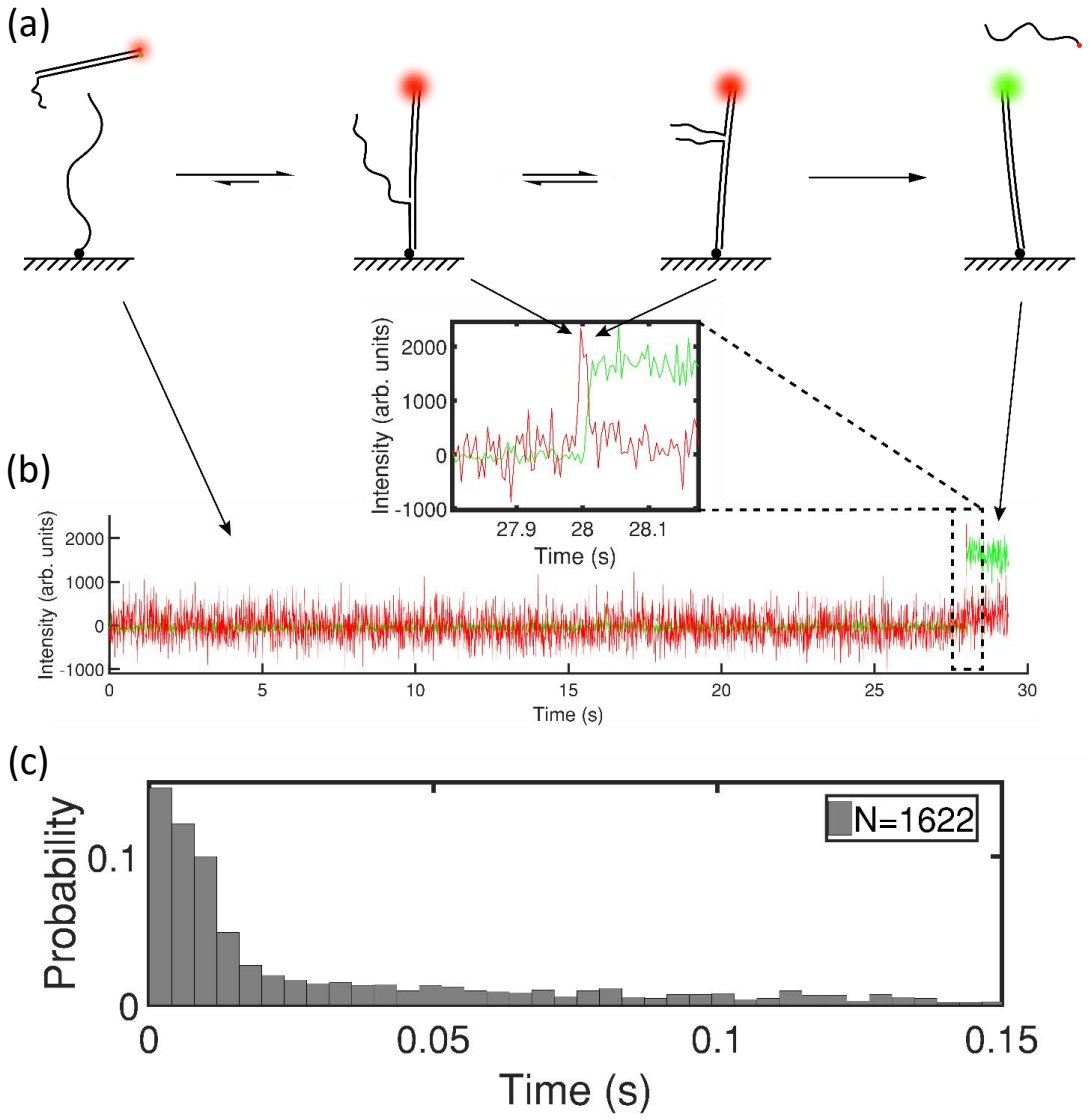


Figure 4.1: Measuring branch migration. (a) Experimental scheme. Invader strands were immobilized on a PEG-passivated cover slip surface via a biotin-NeutrAvidin linker. Partial duplexes were labeled with a FRET pair (Cy5, incumbent; Cy3, substrate) and were flowed into the flow cell. After a diffusion search process, the partial duplex binds to the toehold and the incumbent strand is displaced. (b) Sample acceptor and donor time traces. Acceptor (Cy5) signal increases upon toehold binding, and high FRET signal is sustained until displacement occurs which is indicated by a low FRET signal. The first passage time is identified as the high FRET lifetime. (c) First passage time distribution. First passage lifetimes are collected for many traces and assembled into a distribution. The binning size is the single frame exposure time ($4\,\mathrm{ms}$). The leftmost bar is the observed probability of donor-only traces.

Table 4.2: Incumbent sequences.

5′-GTCGGAATTTTAAT/Cy5/-3′
5′-/Cy5/ATTAAAATTCCGAC-3′
5′-TGTTCGTACGTGTG/Cy5/-3′
5′-/Cy5/CACACGTACGAACA-3′
5′-TTGGGTGTGTGAAA/Cy5/-3′
5′-/Cy5/TTTCACACACCCAA-3′
5′-AAACTCTCTCTAAA/Cy5/-3′
5′-/Cy5/TTTAGAGAGAGTTT-3′
5′-CTTTTAGGAGGAAA/Cy5/-3′
5′-/Cy5/TTTCCTCCTAAAAG-3′
5′-GAACTATAGTCGTC/Cy5/-3′
5′-/Cy5/GACGACTATAGTTC-3′

Table 4.3: Substrate sequences.

5′-CA/Cy3/ATTAAAATTCCGACAACACCAGGT-3′
5′-ACCTGGTGTTGTCGGAATTTTAAT/Cy3/AC-3′
5′-CA/Cy3/CACACGTACGAACAAACACCAGGT-3′
5′-ACCTGGTGTTTGTTCGTACGTGTG/Cy3/AC-3′
5′-CA/Cy3/TTTCACACACCCAACCCAATAACC-3′
5′-GGTTATTGGGTTGGGTGTGTGAAA/Cy3/AC-3′
5′-CA/Cy3/TTTAGAGAGAGTTTTATTTGATTG-3′
5′-CAATCAAATAAAACTCTCTCTAAA/Cy3/AC-3′
5′-CA/Cy3/TTTCCTCCTAAAAGACCACCACCT-3′
5′-AGGTGGTGGTCTTTTAGGAGGAAA/Cy3/AC-3′
5′-CA/Cy3/GACGACTATAGTTCAACAGCAGGT-3′
5′-ACCTGCTGTTGAACTATAGTCGTC/Cy3/AC-3′

4.3.2 Experimental Setup

Molecules were observed with objective-type total internal reflection fluorescence microscopy with a commercially available microscope (IX81; Olympus). Fluorophores were excited by a $532\,\mathrm{nm}$ laser (BWN-532-50E, B&W Tek). Images were 2×2 binned and captured with an electron-multiplying charge-coupled device (DU-897ECS0-#BV; Andor Technology), and images were recorded at 228 fps with $3.96\,\mathrm{ms}$ exposure time using Micro-Manager software [133]. This high frame rate was achieved by cropping the image height to 64 pixels. Experiments were performed on flow cells constructed as previously described in

Le and Kim [108] while flow volume and flow rate ($900\,\mu\mathrm{L/s}$) was controlled by a syringe pump (NE-1000; New Era Pump System).

The surface was passivated with polyethylene glycol (PEG) to minimize nonspecific binding. After neutravidin coating, the biotin-containing invader molecules were immobilized by flowing in at a concentration of $1\,\mathrm{nM}$. Next, $20\,\mu\mathrm{L}$ of partial duplexes were pumped into the flow cell at $200\,\mathrm{pM}$ in an oxygen-scavenging imaging buffer [109], which contained $1\,\mathrm{mM}$ 6-hydroxy-2,5,7,8- tetramethylchroman-2-carboxylic acid (Trolox), $5\,\mathrm{mM}$ protocatechuic acid, $100\,\mathrm{nM}$ protocatechuate 3,4-dioxygenase, and $100\,\mathrm{mM}$ Tris-HCl (pH 7).

An appearance of high FRET signal signified initiation of branch migration. A low FRET signal appeared as branch migration concluded. The FRET signal time series was recorded and analyzed using in-house MATLAB software. The lifetime of the high FRET state was observed for many molecules to collect a distribution of branch migration times.

Nonlinear least-squares fitting was performed on the cumulative distribution with lsqcurvefit of the MATLAB Optimization Toolbox. DNA duplex free energies were calculated as the arithmetic mean of calculations from oligoprop of the MATLAB Bioinformatics Toolbox [134, 135, 136, 137].

4.4 Results

4.4.1 Branch migration time is measured by FRET

Partial duplexes were constructed by annealing Cy3-labeled substrate molecules and Cy5-labeled incumbent molecules. Invader molecules were biotinylated near the end containing the toehold sequence and immobilized onto the surface (see Figure 4.1(a)). Partial duplexes were flowed into the chamber and the appearance of a high FRET state immediately followed by a low FRET state indicated completion of strand displacement (see Figure 4.1(b)). The first passage time for individual branch migration events was measured from the high FRET state lifetime.

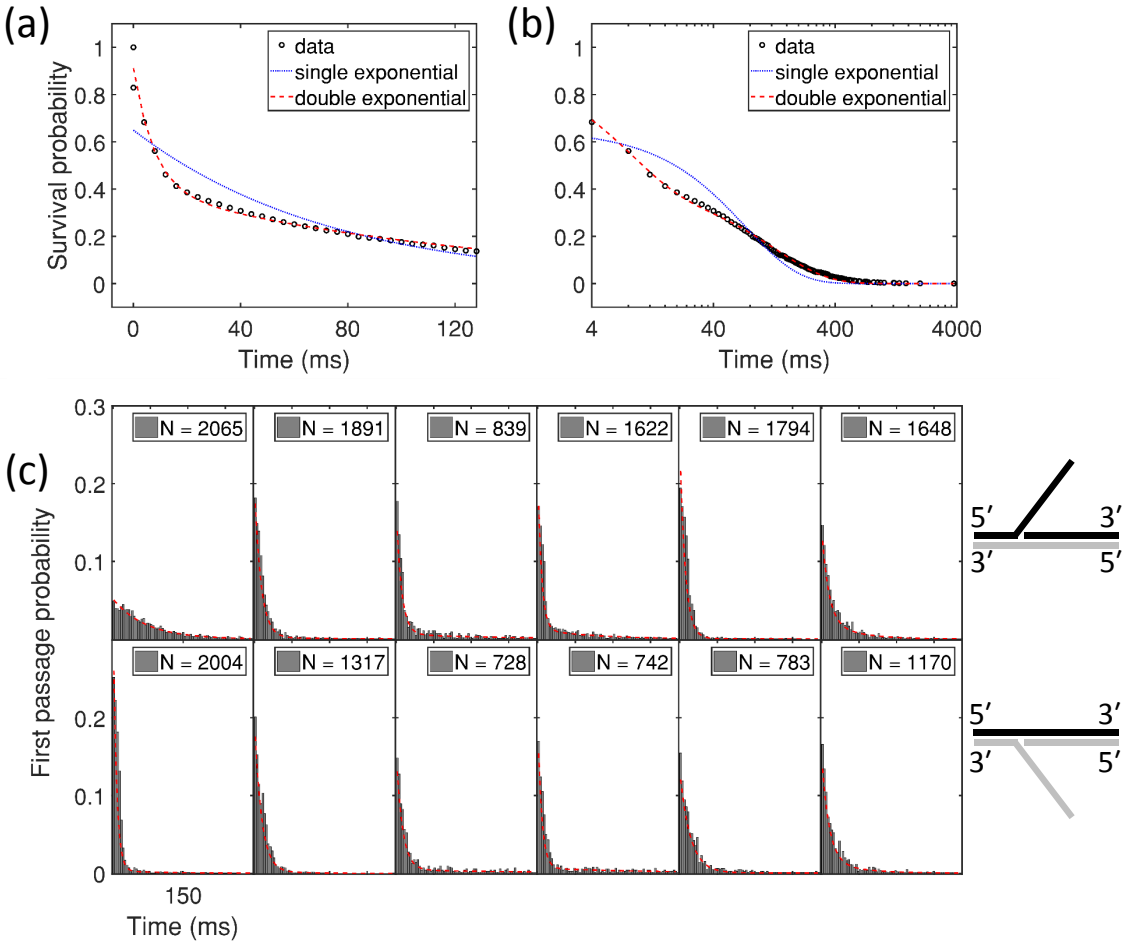


Figure 4.2: Branch migration histograms. (a,b) Exponential fits to survival times. Survival distributions were fitted to single (blue dots) and double (red dash) exponentials. Distributions were better fit by double exponentials. (c) First passage time distributions for different sequences. First passage distributions are plotted for various sequences with equal axes for each plot. Each column corresponds to complementary invaders where the invader and substrate swap roles. The red dash represents the derivative of the double exponential survival curve fit. To the right is a schematic of complementary invaders. Strands with identical sequences share the same color. Therefore, the invader in the topmost system has the same sequence as the substrate in the bottom sequence. Ends are denoted with their respective polarity (5′ or 3′) to help orient the reader.

The distribution of first passage times is constructed from observing hundreds of strand displacement events for 12 strand displacement systems (see Figure 4.2(c)). Each distribution is binned to single exposure times ($3.96\,\mathrm{ms}$). The first bar in each histogram represents our estimate of events occurring faster than a single exposure time. We can estimate this number by counting the number of donor-only traces that contain only low FRET states with no preceding high FRET state. Besides fast displacement, there are two other potential sources of donor-only traces: fluorescently inactive Cy5 and single strand substrates. We performed control experiments with a truncated invader strand containing only the toehold sequence and found fewer than 10% of the spots to be donor-only. We measured the effect of salt concentration for several sequences, and, interestingly, the branch migration time showed very little dependence on cation species and salt concentration.

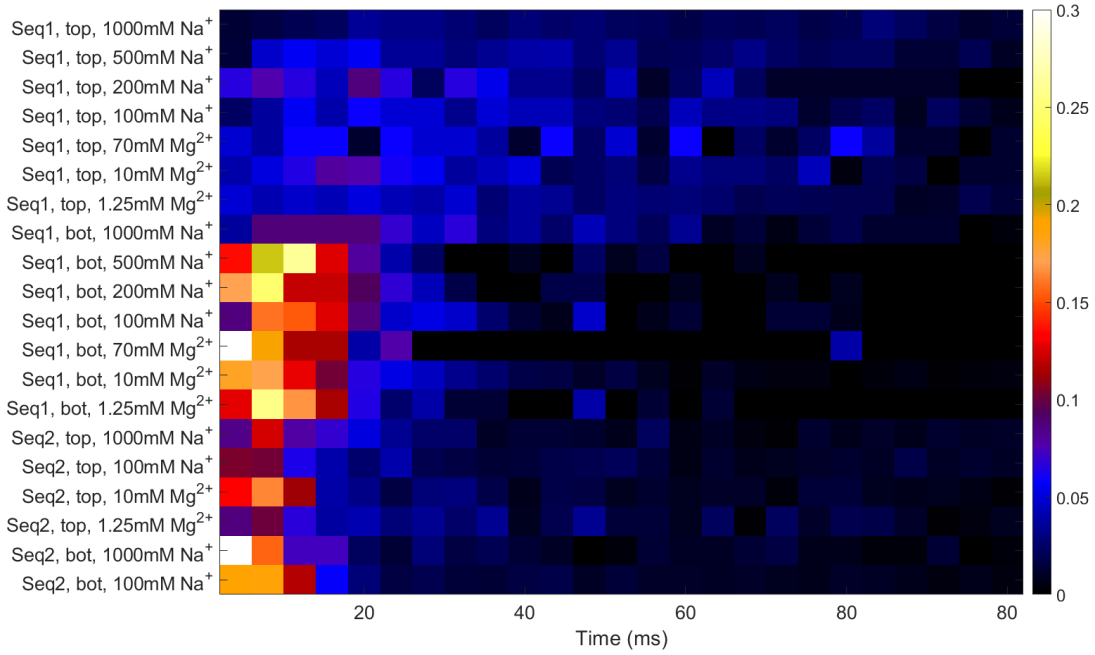


Figure 4.3: Salt dependence. We measured the branch migration times for various sequences under several salt conditions and present their histrograms as a heat map with color representing probability. The sequence numbers correspond to the first two columns in Figure 4.2 from left to right. The 'top' and 'bot' designations refer to their corresponding row in Figure 4.2. These experiments were performed under an $80\times$ excess of incumbent strands to substrate strand.

In order to further quantify the distributions, we attempted single and double exponential fits to the survival probabilities which were calculated as the cumulative sum of the measured waiting time distributions (see Figures 4.2A,B). The survival curves are much better fit by the two-term exponential. The first passage time distributions are presented with their corresponding fit which was calculated from the derivative of the double exponential survival curve fit (see Figure 4.2(c)).

4.4.2 Complementary Invasion

To gain better insight into the kinetics of two-way branch migration, we examined the effects of the "side" of invasion on the first passage time distribution. To accomplish this, we studied systems with complementary invaders which swap the roles between the invader and substrate such that the new invader was the old substrate and *vice versa*, while the incumbent was replaced with its reverse complement (see Figure 4.2(c), right). Figure 4.2(c) is arranged so that columns of distributions represent complementary invasion presenting distributions from invaders with a $5'$ proximal toehold in the top row and invaders with a $3'$ proximal toehold in the bottom row. We performed this experiment with 6 pairs of branch migration systems for a total of 12 sets of corresponding invaders, substrates, and incumbents. We noticed an observable shift in the first passage time distribution through complementary invasion (see Figure 4.4) particularly for the pair of systems in the leftmost column of Figure 4.2(c).

In an effort to better understand the physical origins of the observed disparity between distributions, we found it helpful to consider the mean first passage time (MFPT) for all DNA sequences considered. We repeated histogram collection 3 entirely separate times for each of the 12 systems. We noticed high sensitivity to the MFPT in very rare, very long-lived events. To account for this, we calculated the MFPT for all data below the 95^{th} percentile. Figure 4.4 shows a clear separation between most complementary invader times. Invaders with a $5'$ proximal toehold are slower than their complementary invaders

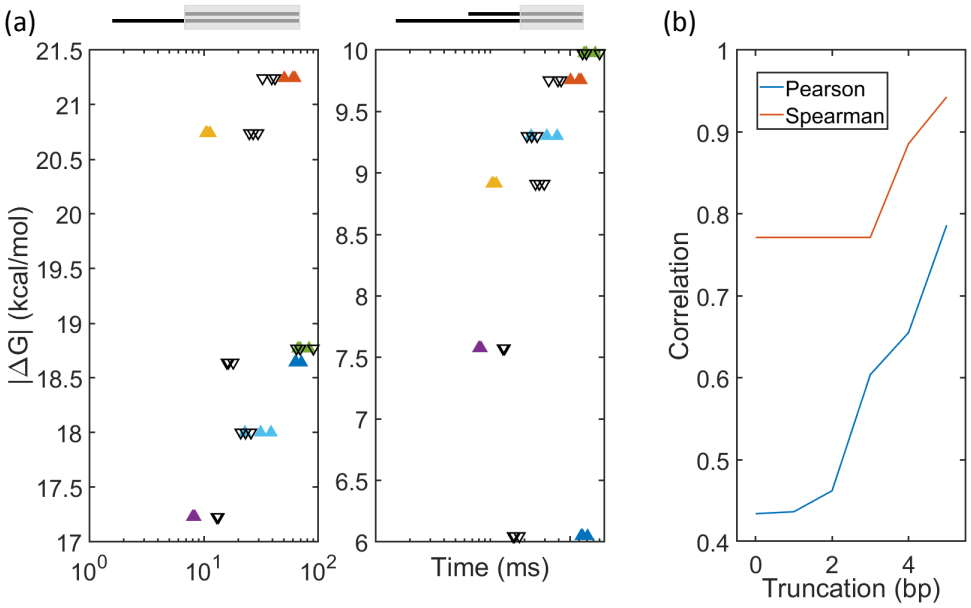


Figure 4.4: Partial duplex stability vs mean first passage times. (a) The mean first passage time is calculated from data below the 95$^{\text{th}}$ percentile. The data collection process was entirely repeated in triplicate for each system considered. Upward facing triangles represent mean times for invaders with a $5'$ toehold while downward facing triangles represent invaders with a $3'$ toehold. The free energy of annealing between incumbent and substrate is calculated for two truncations (left, 14bp; right, 7bp). Strong correlations appear between the mean first passage time and stability of the final 7 base pairs of the partial duplex. (b) Pearson and Spearman correlation coefficients are calculated for free energies vs $3'$ invader toehold log-times. Both correlations clearly increase with an increasing number of truncated proximal base pairs.

with a $3'$ proximal toehold in three systems, faster for two systems, and roughly equivalent in one system. Further, we calculated the interaction strength (absolute duplex free energy) between substrate and incumbent base pairs and plotted it against the MFPT (see Figure 4.4). Complementary invader pairs are represented with upward ($5'$ invader toehold) and downward ($3'$ invader toehold) facing triangles. We observed little correlation when considering the entire duplex sequence (see Figure 4.4, left); however, by truncating the region of consideration to the toehold distal half of the partial duplex (7bp), a clear trend emerges showing exponential dependence between the time and duplex stability (see Figure 4.4, right).

4.4.3 RNA Invaders

Spurred by an interest in RNA-DNA interactions, we replaced the DNA invader for the sequence in the leftmost column of Figure 4.2(c) with an RNA invader while maintaining the incumbent and substrate molecules as DNA. We measured the branch migration time for both sides of invasion and again found that branch migration time is polar; however, the polarity reversed meaning faster/slower sides of invasion switched (see Figure 4.5). In other words, the faster direction for DNA invasion became the slower direction for RNA invasion and *vice versa*.

4.4.4 Loaded Invaders

As a further exploration of the effect of dangling ends, we performed experiments with invaders whose branch migration domain was extended with 5 thymine nucleotides at the toehold distal end. We noticed a significant increase in first passage time which even included an initial lag (see Figure 4.6). Without an analytical description, we used a Gillespie algorithm to simulate a branch migration process, whose forward or reverse step rate depends on the length of the dangling end. In our modified 1D random walk simulation, a strand with a longer dangling end would have a geometrically slower rate. Our simulation

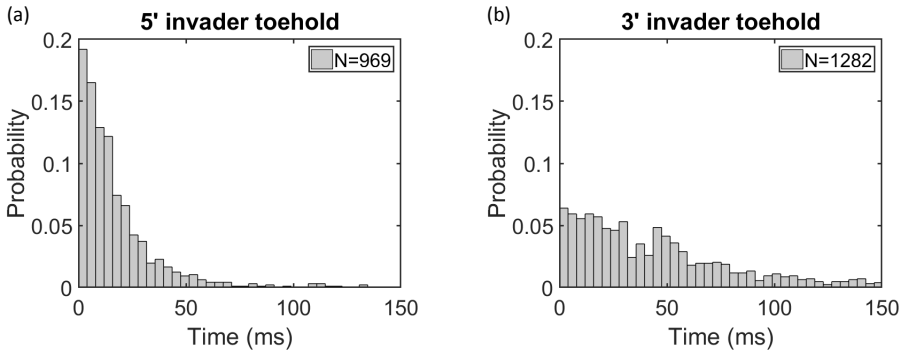


Figure 4.5: RNA invaders with DNA partial duplex. We measured branch migration times for RNA versions of the invaders from the first column of Figure 4.2 where the direction from the top and bottom histograms correspond to the directions in (a) and (b) respectively. Again, we found a difference in passage time depending on the side of invasion. Further, we noticed that the relative times switched with respect to the side of invasion.

matched key features of the data, including lag and long tail, with reasonable parameters for the step rate ($200\,\mathrm{s}^{-1}$) and geometric increase factor (1.05).

4.5 Discussion

Using a FRET-based detection scheme, we separated the timescales of the two main steps of strand displacement: concentration-dependent, bimolecular toehold formation and unimolecular branch migration. We directly observed the first passage time of two-way branch migration and found very little change in the branch migration time as we varied salt concentration for several sequences Figure 4.3. This is not an obvious result as a single step of two-way branch migration is thought to be preceded by many fraying/zippering events at the junction [25], implying that even a single branch migration step is a complicated multi-step process. Given the observed lack of salt dependence, it is reasonable to propose that branch migration kinetics are dominated by salt independent processes, such as fraying/melting [119].

We also found the waiting time distributions to be best fit by a two-term exponential curve. This is very surprising because any kinetic model built entirely from sequential

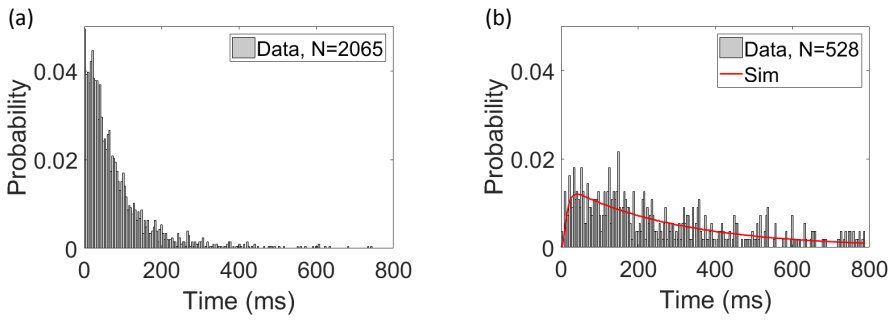


Figure 4.6: Loaded invader. (a) The histogram from the top left corner of Figure 4.2 is recapitulated here. (b) We appended the invader from (a) with 5 thymine residues at the toehold distal end and recorded 528 branch migration events. The overall time slowed down dramatically and a lag became noticeable. The solid curve is a simulation described in the text.

neighboring steps will always produce a noticeable lag or will become single exponential for step times much faster than the binning time. However, our observations of nonmono-exponential behavior are not unique as previous studies of DNA dissociation have found distributions best described with two-term [138, 139, 140, 141], multi [142], or stretched exponentials [143]. The reasons provided for such behavior include dispersion in DNA sample [139], surface heterogeneity [140], and presence of multiple pathways [143]. We cannot ignore the possibility of dispersion in HPLC-purified, commercial oligos; however, any such defect in synthesis would have to be consistently found across all 14 systems studied (38 total oligos) each with very different sequences. Further, surface interactions are also possible, but at $1\,\mathrm{nM}$ invader concentration, the average density is $\sim 1/(120\,\mathrm{nm})^2$ which is far beyond the contour length of the invaders. While these arguments do not completely rule out instrumental effects, they do invite discussion of possible two-way branch migration mechanisms that have not been previously considered.

A fundamental component of strand displacement is duplex disruption and dissociation. While there are few studies of single-molecule two-way branch migration, plenty of attention has been paid to simple duplex dissociation [144, 145, 146, 147, 143, 140, 139, 148, 138, 141]. Consider the naive model of unzipping at the ends which is, essentially, a

1D lattice with reflecting boundaries where two random walkers are allowed to annihilate each other upon their encounter. The analytical solution for the encounter time has been shown to be an infinite sum of exponential distributions [149] with the full sum predicting a lag or single exponential distribution for observation times much longer than single step times [144]. It is possible that our experimental timescale prevents us from observing more than the first two leading terms.

We have further demonstrated that side of invasion produces a very significant change in the observed distribution and MFPT. This is surprising due to the high degree of symmetry between complementary invasion. For example, consider any fully-zipped "snapshot" of branch migration (see Figure 4.1(a)). In each snapshot, the identical sequence of base pairs between the invader/incumbent strands and the substrate strand are preserved. Moreover, this exact sequence is preserved for complementary invaders. In other words, regardless of side of invasion, the exact same base pair must be broken and formed along every corresponding step of branch migration in both systems. Despite this tremendous symmetry between corresponding systems, branch migration times for conjugate sides are not identical. In light of this, there are two obvious breaks in symmetry: base-flipping and dangling ends.

The existence of dangling ends distinguishes individual steps of branch migration as the dangling ends of both the invader and incumbent have distinct sequences at every step. This is true even in the simplest case where the invader and incumbent are homopolymers. In this simple case, while the sequences of the dangling ends for the invader and incumbent can be identically matched in pairs, the polarity of the dangling ends are always reversed between the invader and incumbent. Further, When considering 2 different sides of invasion, the dangling ends are again the source of symmetry breaking between the two sets of invaders and incumbents; furthermore, when comparing different sides of invasion, both the sequence and the polarity of the dangling ends change to the reverse complement.

Single DNA nucleotide dangling ends have been shown to alter the free energy of nu-

cleic acids on the order of Watson-Crick base pairing ($\sim \pm 1\,\mathrm{kcal/mol}$) [150]. Generally, these observed free energies confer stability to $5'$ dangling ends over $3'$ dangling ends for DNA duplexes. This suggests that a DNA strand invading from $3'$ to $5'$ will have a more stable dangling end than the incumbent that it is challenging. Likewise, a complementary DNA strand invading from $5'$ to $3'$ will have a less stable dangling end than the incumbent that it is challenging. Hence, we predict that branch migration would typically be faster for $3'$-to-$5'$ invasion, which is supported by our data (see Figure 4.4).

Like DNA, our measurements of RNA show the existence of a faster direction of branch migration. Further, directional preference of toehold-mediated RNA strand displacement has been shown to exist in simulations [132]. Interestingly, this directional preference is reversed for an RNA invader. This reversal is also in agreement with thermodynamic measurements that show $3'$-dangling end conferring equal or greater stability to RNA duplexes than DNA counterparts [150, 151, 152]. Moreover, three dimensional structures of various RNA junctions show markedly more frequent stacking of the $3'$-dangling base with the adjacent base pair than the $5'$-dangling base [153]. Thus, the RNA invader forms stronger base pairs with the substrate when it invades from the $3'$-end of the incumbent, which results in a faster displacement rate. It is also possible that the one or two pre-stacked bases in the $3'$-dangling end of the invader can zipper with the substrate strand at a faster rate [153]. The different effects between RNA and DNA invader likely arise from the different duplex geometries. RNA-DNA duplex has an A-form helix which prefers stacking of the $3'$ dangling end. In comparison, DNA-DNA duplex might prefer stacking of the $5'$ dangling end due to the B-form helix. With the rich diversity found in DNA-DNA and RNA-DNA branch migration, it will be interesting to test xeno-nucleic acids (XNA) whose structures extend beyond canonical A- and B-forms [154]. These results have exciting implications for biology. Recently, it was shown that transcription is regulated by the RNA-DNA/DNA-DNA upstream edge (rear end) of the transcription bubble [43]. The authors found the RNA-DNA interaction inhibits forward translocation of the RNA poly-

merase (RNAP) while DNA-DNA base pairing inhibits reverse translocation. The polarity of these junctions is such that the RNA has a $5'$ dangling end. Our results are consistent with these findings as we would predict an RNA with a $5'$ dangling end to be a weaker invader which would inhibit transcription but, ultimately, bias the direction downstream. Many other biological systems involve protein mediated strand exchange. Therefore, it is conceivable that systems such as homologous recombination and CRISPR/Cas could present directional preference.

As noted earlier, dangling ends are not the only source of symmetry breaking between complementary invaders. Conceivably, base flipping for individual nucleotides could be different for each base and, therefore, account for differences between complementary invaders. However, a previous study compared base flipping rates between RNA and DNA which were found to be the same [155]. This implies that even if base flipping contributed to differences between complementary invaders, it still cannot account for the observed polarity reversal between RNA and DNA.

We also measured the effect of increasing the length of the invader branch migration domain and found a significant increase in the first passage time. This is in agreement with a previous study that explained changes in transcription mechanics as a result of increases in drag due to the presence of nascent RNA in the RNA complex [156]. Therefore, it is conceivable that we could explain our observed slow down as an increase in resistance due to the longer dangling end. Indeed, as shown in Figure 4.6, we can obtain similar first passage time distributions when simulating such a process.

In Chapter 3, we posited a three-step model for toehold-mediated strand displacement which began with toehold formation and proceeded through branch migration until the incumbent spontaneously dissociated [94]. To explore the hypothesis of spontaneous dissociation, we considered the interaction free energy between the toehold distal base pairs of the partial duplex. As we truncated the region of interest, we found a significant correlation between the measured MFPT and calculated duplex free energy. In accordance with the

three-step model, the MFPT increased with increasing interaction strength. In other words, stable incumbent base pairs distal to the toehold take longer to displace. Further, such a correlation implies incumbent dissociation is the rate limiting step. However, this is not consistent with the three-step model as the original model included the presence of a mismatch to act as a barrier to branch migration. Without a kinetic barrier, there is no reason for the system to cease branch migration and wait for a (slower) spontaneous dissociation process.

Without a consistent interpretation from the three-step model, it is worthwhile to consider alternative mechanisms of displacement. A similarity can be found in a symmetric random random walk in which rates between neighboring lattice sites are equal but independent and identically distributed with rates between other lattice sites [157, 158]. In this Sinai model, the later steps contribute more to MFPT than the earlier steps. This is consistent with our observed correlation as incumbent dissociation must be the last step of such a hypothetical process.

Alternatively, competitor-induced dissociation has been shown to be a viable mechanism at very high competitor concentrations ($> 20\,\mathrm{uM}$) [126]. Dissociation is rapidly accelerated as the competitor occludes the rapid rebinding of the two original binding partners. A toehold bound invader would certainly have a high effective concentration relative to the partial duplex ($\sim 1\,\mathrm{mM}$). Such a mechanism is also consistent with the observed correlation. This model would also retain other key features expected from branch migration such as an increase in waiting time with an increase in length of the partial duplex [25]. However, the survival curves measured from competitor-induced DNA dissociation were found to be single exponential.

4.6 Conclusion

Our FRET-based strand displacement assay allows investigation of branch migration post toehold formation. Surprisingly, we found that such distributions were better described by

two-term exponentials. We showed that branch migration time is dependent on the side of invasion to which we hypothesize dangling ends as the source of dissimilarity. Additionally, we showed that mean first passage times strongly correlates with the stability of the duplex region distal from the toehold which suggests direct dissociation as a plausible kinetic pathway. We hope our work will inform the communities of DNA nanotechnology and biology with a better understanding of two-way branch migration.

CHAPTER 5
BRANCH MIGRATION WITH A SINGLE FLUOROPHORE

5.1 Introduction

As explained in Chapter 4, the timescales between bimolecular toehold formation and unimolecular branch migration can be decoupled from each other using a FRET based scheme. This scheme has allowed novel exploration of the branch migration process and provided many exciting observations; however, there are still many exciting questions to be answered about branch migration surrounding the topics of length dependence, individual step rate, and sequence dependence. Unfortunately, such a scheme as described in Chapter 4 requires modifications of all three participants of DNA strand displacement which can be quite costly when considering the large scale of experiments needed to answer questions regarding length and sequence dependence. Therefore, in this chapter, I will discuss a new, low-cost scheme that requires a single fluorophore to measure branch migration. I will discuss preliminary results related to length and single nucleotide polymorphism (SNP) location dependence. Finally, I will discuss future prospects and potential complications of the single fluorophore method.

The cost of DNA sequencing has decreased exponentially over the past three decades (see the blue curve in Figure 5.1(a)). In particular, the price of DNA sequencing has even outpaced Moore's law. As the price continues to drop, the dream of personalized medicine, the $\$1000$ genome, is almost within reach [159]. In comparison to whole genome synthesis and sequencing cost, the price of custom oligonucleotide synthesis has remained relatively stagnant with only modest decreases over time. Moreover, bespoke DNA oligos with covalently attached modifications can cost over an order of magnitude more than the same strand without modifications (see Figure 5.1(b)). Given the huge price increase as-

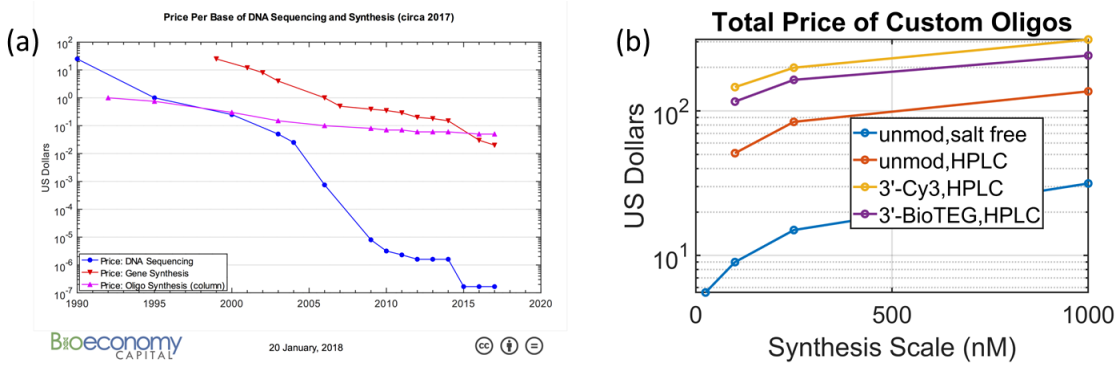


Figure 5.1: Oligo prices. (a) The price of reading (blue) and writing (red) genomic DNA has exponentially decreased over the past three decades with DNA sequencing outpacing Moore's law. By comparison, DNA oligonucleotide synthesis (magenta) has seen a much slower decline in price. This figure is taken from reference [27] and shared under a CC BY-ND license. (b) The prices of DNA oligo tube synthesis for a random sequence (GCG GGC AAT ATG TAC) are plotted for various modifications (unmodified, 3'-Cy3, and 3'BioTEG), purification methods (salt free and high-performance liquid chromatography), and synthesis scales ($25\,\mathrm{nM} - 1\,\mu\mathrm{M}$). Modified DNA cost can be more than an order of magnitude higher in cost than unmodified DNA. The prices were obtained from the website of Integrated DNA Technologies, Inc. (IDT).

sociated with modified DNA oligos, it is impractical to conduct large scale sequence and length dependence studies of branch migration with the FRET scheme. Therefore, in order to accomplish our experimental goals, we must use a method with unmodified invaders, incumbents, and substrates.

While we require the strand displacement participants to be unmodified, our method must include a biotinylated attachment to the surface as well as a fluorescent reporter. Given our success with the FRET method, we decided to again immobilize the invader molecule on the surface. With these requirements in mind, the main idea behind our method is to contain the modified oligos to preserved sequences that can hybridize to the unmodified invader, incumbent, and substrate molecules (see Figure 5.2). A biotinylated anchor molecule attaches to the surface and hybridizes to an invader molecule. The sequence is short ($\sim 10\,\mathrm{bp}$) and kept constant over the course of all possible experiments. The incumbent is designed with a fixed overhang sequence that can hybridize with a fluorescent indicator probe. Although not covalently attached to the substrate, the fluorescent indicator

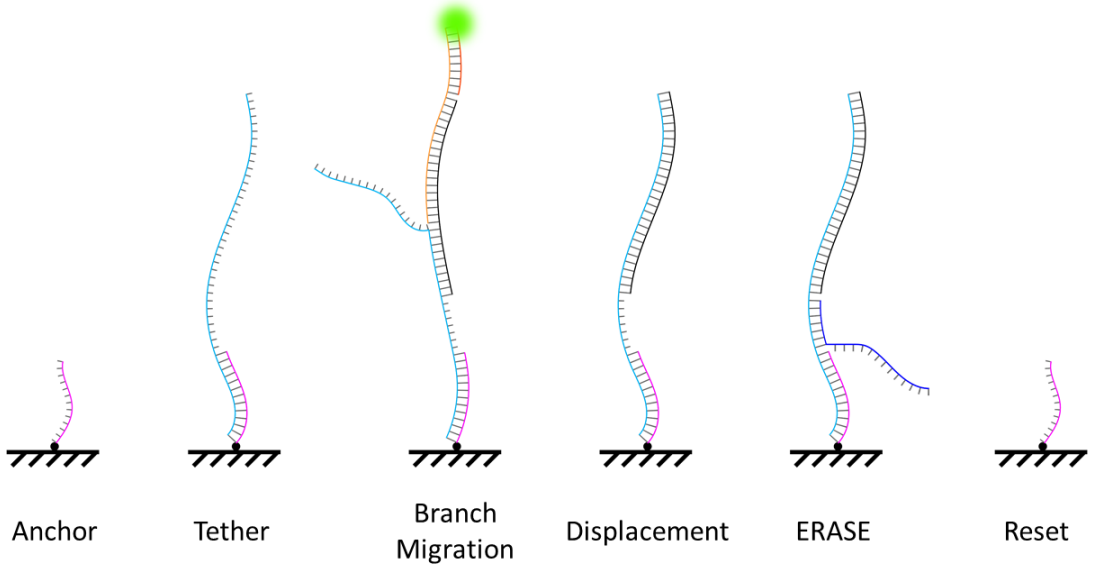


Figure 5.2: Single fluorophore schematic. A fixed-sequence, biotinylated anchor molecule is attached to the surface and hybridizes to the invader molecule which acts as a tether. The fluorescently labeled partial duplex consists of the substrate, an incumbent with a fixed overhang sequence, and a fluorescent indicator probe which can anneal to the overhang sequence. Branch migration is measured by the fluorescent signal dwell time. After strand displacement is complete, ERASE can be used to release the duplex and reset the surface.

probe is stabilized on the partial duplex by coaxial stacking. The lifetime of the fluorescent signal will correspond to branch migration passage time. ERASE is incorporated into this method so that the surface can be reset upon completion of strand displacement.

5.2 Methods

The system was designed using mfold [115] to minimize the possibility of secondary structure interactions. Custom DNA oligomers were purchased from Integrated DNA Technologies. The 13-nt anchor molecule was biotinylated at the 3′ end. The 10-nt fluorescent indicator molecule was labeled with Cy3 at the 5′ end. The switch molecule which ERASEs the surface is unmodified and is 20-nt in length with a 7-nt toehold near its 3′ end. The invader, substrate, and incumbent molecules were unmodified. All modified oligos were

Table 5.1: Fixed sequence oligos.

Anchor	5′-TCAATTCGTCGTC/BioTEG/-3′
Indicator probe	5′-/Cy3/TCTCGACGTC-3′
Switch	5′-TTTCACTTCAATTCGTCGTC-3′

Table 5.2: Variable sequence oligos. The variable branch migration domain is represented by X. X^* represents the reverse complement.

Invader	5′-GACGACGAATTGAAGTGAAAGATGTAGAGC/X/-3′
Incumbent	5′-/X/GACGTCGAGA-3′
Substrate	5′-/X^*/GCTCTACATC-3′

HPLC purified, and unmodified oligos were purified under salt-free conditions. To distinguish from sequence dependent effects, the branch migration region was chosen to be a homopolymer. The specific sequences are presented in Tables 5.1, 5.2, and 5.3.

Partial duplexes were constructed by combining substrate and incumbent at a 1:1 ratio of $20\,\mu\mathrm{M}$ each in $1\,\mathrm{M}$ NaCl and $10\,\mathrm{mM}$ Tris at pH 7. The mixture was heated to $95\,^{\circ}\mathrm{C}$ and slowly cooled for 3 hours to $4\,^{\circ}\mathrm{C}$. After this, indicator probe was added to a final concentration of $20\,\mu\mathrm{M}$ so that the final proportion of incumbent, substrate, and indicator was 1:1:1.

PEG-passivated flow cells were prepared as described in Chapter 4. We introduced $20\,\mu\mathrm{L}$ of anchor molecules at $20\,\mathrm{nM}$ and waited for $2\,\mathrm{min}$. We then introduced $20\,\mu\mathrm{L}$ of $50\,\mathrm{nM}$ invader molecules and waited for $5\,\mathrm{min}$. The experiment began as we introduced $20\,\mu\mathrm{L}$ of $1\,\mathrm{nM}$ partial duplex-indicator probe complex. After recording the data, the surface was reset by introducing $20\,\mu\mathrm{L}$ of $500\,\mathrm{nM}$ switch molecule.

Data was acquired as described in Chapter 4. Since the indicator probe was labeled with a Cy3 fluorophore, only the donor channel was observed. The low partial duplex concentrations helps reduce background signal but also causes strand displacement events to remain rare. As such, multiple strand displacement events rarely appear simultaneously in a frame. The raw data of a single frame in Figure 5.3 presents the rare event of two simultaneous strand displacement events. As non-specific interactions between the partial duplex

Table 5.3: Branch migration domain sequences.

Invader and Incumbent, X
5′-AAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAAAAAAAAAAA-3′
5′-GAAAAAAAAAAAAAAAAAAAAAAAAA-3′
5′-AAGAAAAAAAAAAAAAAAAAAAAAAA-3′
5′-AAAAGAAAAAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAGAAAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAGAAAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAGAAAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAGAAAAAAAAAAAAA-3′
5′-AAAAAAAAAAAAAAGAAAAAAAAAAA-3′
Substrate, X*
5′-TTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTC-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTT-3′
5′-TTTTTTTTTTTTTTTTTTTTTTTTCTTTTTTTTTTT-3′

and surface can appear as false positive branch migration events, control experiments were performed for each surface before data collection. Partial duplex must be introduced into the flow cell and non-specific interactions must be observed before each round of surface modification (neutravidin, anhor, and invader). The experiment can only proceed if very few surface interactions are observed.

MATLAB was used to analyze the data. Candidate spots were identified as peaks detected by thresholding each frame individually. Spot candidates were cross-checked for overlapping positions, and time traces were generated from the surviving candidates. Branch migration time is inferred from fluorescence lifetime.

5.3 Preliminary Results

While we have begun to understand DNA branch migration in Chapter 4, there are still many open questions. Can we observe length dependence as predicted by the 1D model? What are the sequence dependent effects of branch migration? Even more fundamental, what is the step rate of branch migration?

To answer these questions, we used TIRF microscopy and a single fluorophore construct to measure branch migration. The partial duplex was hybridized with a fluorescent probe which served as an indicator for branch migration dwell time. Histograms were constructed from hundreds of dwell time measurements. We found these histograms to be fit well to a single exponential (see Figure 5.4).

5.3.1 Length dependence

We sought to test the 1D model's prediction of length dependence; however, as shown in Chapter 4, branch migration times are highly dependent on sequence as well as direction of invasion. In order to distinguish length effects from sequence effects, we designed the branch migration region to be a homopolymer, poly(A) for the invader/incumbent and poly(T) for the substrate. We measured branch migration histograms for lengths ranging

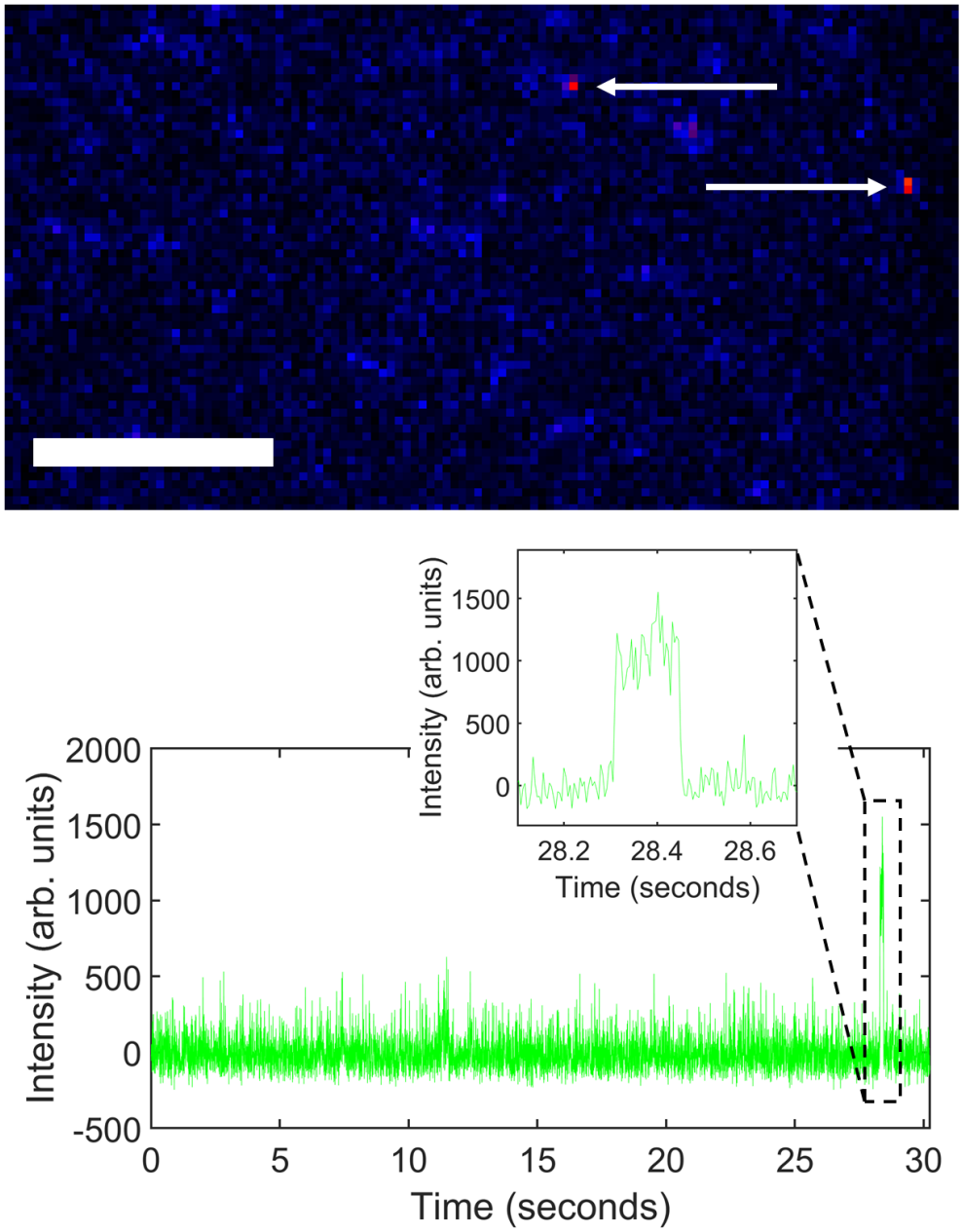


Figure 5.3: Raw fluorescence data. Above. A single frame of data shows two simultaneous branch migration events (arrows). The scale bar is $10\,\mu\mathrm{m}$. Below. After detecting spots, time traces are constructed by summing the 3×3 neighborhood of the peak location. The zoomed inset shows the branch migration event lasting for $\sim 150\,\mathrm{ms}$.

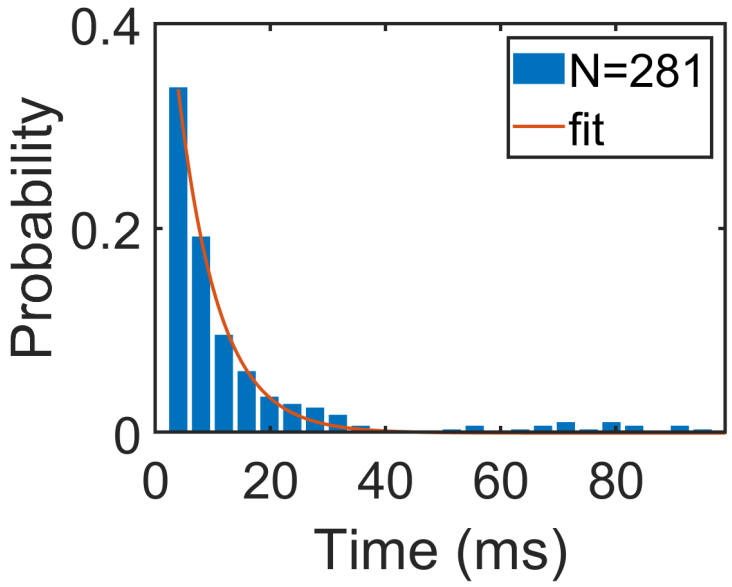


Figure 5.4: First passage histogram with fit. The first passage time histograms for the 15nt ($N = 281$) branch migration region is fitted by a single exponential.

from 15nt to 30nt in 3nt steps for a total of 6 different systems (see Figure 5.5). The mean times show an increasing trend with increasing length. We fit the data to the 1D model prediction of lattice length dependence,

$$\langle \tau \rangle = \frac{n(n+1)}{2k_{step}} \tag{5.1}$$

where n is the number of lattice sites and k_{step} is the single step rate of branch migration. The one parameter fit yields an branch migration single step rate of $k_{step} = 4000\,\mathrm{s}^{-1}$ which corresponds to $\tau_{step} = 250\,\mu\mathrm{s}$.

5.3.2 SNP dependence

After estimating the branch migration single step rate, we turn to measuring the effect of a point mutation and the dependence of location on branch migration. We chose a single nucleotide polymorphism (SNP) that exchanges purine for purine ($A \rightarrow G$) and pyrimidine for pyrimidine ($T \rightarrow C$). Due to the higher free energy costs associated by breaking a G:C

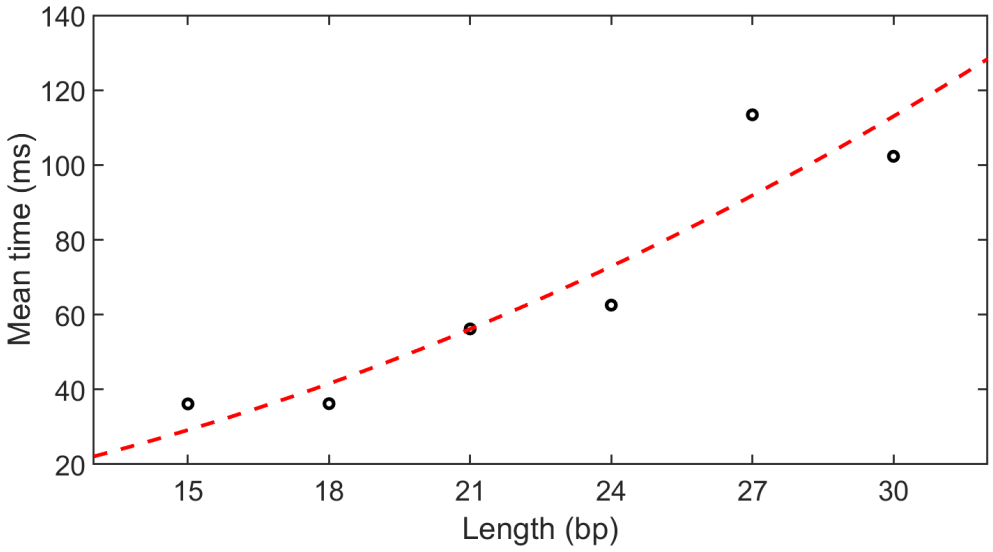


Figure 5.5: Length dependence. The branch migration region length is increased in 3nt increments from 15nt to 30nt. The clear increase in the data (circles) can be fitted (red dash) by a 1D model (see Equation 5.1) and a step rate can be extracted.

bond over an A:T bond, we expect that the branch migration rate should increase with the presence of a SNP. Further, the 1D model predicts a linear increase in mean time with increasing SNP position.

$$\langle \tau \rangle = \sum \frac{i}{k_i} = \sum_{i \neq s} \frac{i}{k_A} + \frac{s}{k_G} \tag{5.2}$$

where k_i is the single step rate at the ith lattice site, s is the location of SNP, and k_A and k_G are the step rates for base pairs involving A and G respectively. We measure the position dependence and found a fit of $k_G = 100\,\mathrm{s}^{-1}$ that corresponds to $\tau_G = 10\,\mathrm{ms}$. This is an order of magnitude higher than k_A obtained from length dependence.

5.4 Future outlook

This method has been demonstrated to measure branch migration time and could be used as a method for large-scale sequence and length dependence study. However, this method does have drawbacks compared to FRET, and there are complications that must be consid-

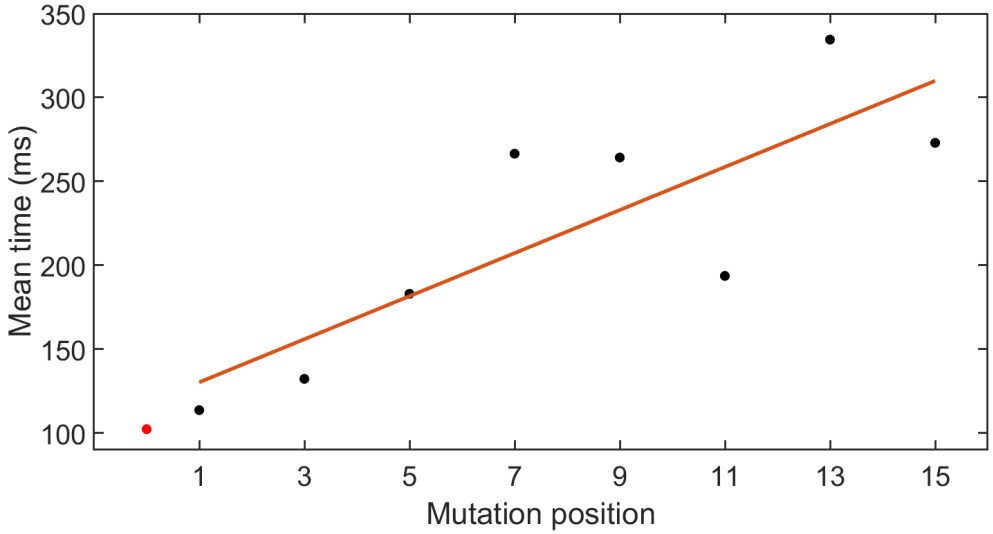


Figure 5.6: SNP dependence. The position of a single A→G mutation was varied in the branch migration region. The red dot is the mean time of the perfect polyA with zero mutation. Numbering begins with position '1' corresponding to the base pair proximal to the toehold with position '15' corresponding to the middle base pair. A clear increase in mean time is observed which is consistent with the intuition of a slower step rate for G:C pairs.

ered when performing these experiments. To abandon the FRET experiment is to abandon the zeroth time bin as this method uses only one fluorophore. Further, the FRET experiment offers a built-in control for discriminating non-specific surface interactions. For poorly prepared surfaces and partial duplex constructs, non-specific interactions drastically increase which can degrade the data as these interactions exist on a similar time scale as branch migration. As such, controls on clean surfaces must be performed before each experiment to check against non-specific interactions. Finally, this experiment uses a homopolymer for the branch migration region. It can be difficult to properly anneal two homopolymers as frame shifts can occur which effectively shorten the branch migration region. Also, a homopurine-homopyrimidine DNA duplex can form a triplex due to mirror symmetry [160, 161].

For its concerns, this method also shows promise as the inclusion of ERASE could be used to multiplex the data acquisition process. The substrate could be extended and

annealed with a fluorescent probe in a manner similar to the incumbent. The experiment would proceed as described except that the partial duplex buffer would contain many different molecular species that each have a different branch migration region. The toehold of the switch could then be used to encode the particular branch migration region. After ERASE, disappearance of that fluorescent signal after a specific switch molecule will allow the experimenter to identify the branch migration region for that trace.

It is possible to conceive alternative schemes that incorporate both FRET and unmodified invaders, incumbents, and substrates. For instance, the substrate could be extended on the end distal to the toehold with a fixed sequence so that another fluorescent probe can anneal. Further, two color schemes that do not use FRET could be possible. In one such scheme, the partial duplex would be constructed identically as in Figure 5.2 yet immobilized on the surface by an anchor molecule. The invader molecule could be extended with a fixed sequence to hybridize to a Cy5 probe. In this scheme, both lasers would be employed so that both channels could monitor the binding and unbinding of labeled molecules to the surface.

CHAPTER 6

EPILOGUE

This thesis is centered on the relationship between single-molecule biophysics and toehold-mediated strand displacement. In Chapter 2, I have demonstrated a practical application of strand displacement in the form of ERASE. In Chapter 3, I explored the mismatch position-dependence kinetics of strand displacement and demonstrated the plausibility of direct dissociation a pathway for strand displacement under the presence of a mismatch. Further, in Chapter 4, I explored the kinetics of branch migration and found a rich dependence on sequence as well as a surprising dependence between complementary sides of invasion which pointed to dangling ends as a cause. Finally, in Chapter 5, I demonstrated the feasibility of a more low-cost method to allow large-scale experimental studies of sequence and length dependence. With this method, a proper exploration of sequence can determine nearest-neighbor parameters and provide an accurate model for branch migration.

Appendices

APPENDIX A

END INVASION AND CONSECUTIVE MISMATCHES

One hypothesis proposed in Chapter 3 was the three-step model which considered strand displacement as a toehold-binding step in equilibrium with branch migration. Branch migration proceeds until halted by the presence of a mismatch, and displacement occurs through direct dissociation of the incumbent. We noted the increased stability, as measured through off-rates, of a duplex compared to a partial duplex that shares the same number of base pairs. We proposed the presence of the invader somehow destabilizes the incumbent which accounts for the increased dissociation rates. One such mechanism is steric hindrance at the junction which prevents the incumbent from fully zipping up along the substrate. Another possibility is competitor-induced dissociation proposed by Gelles *et al.* [126] where the dissociation of a complex is rapidly accelerated in the presence of high competitor concentrations. To explore other possible mechanism of DNA strand displacement, we used the fluorescence recovery scheme described in Chapter 3.

Using the methods developed in Chapter 3, we first explored direct dissociation of the incumbent by measuring strand displacement for several variations of the invader with a first position mismatch (see Figure A.1). We first tested 4 different variations of branch migration region: no branch migration region, entirely mismatched branch migration region,

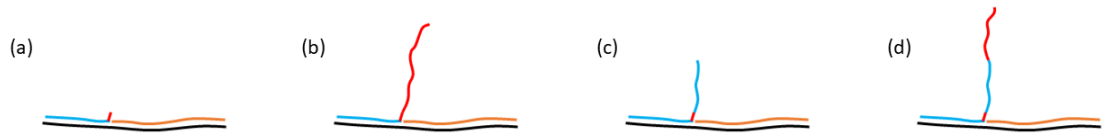


Figure A.1: Testing spontaneous dissociation. No strand displacement was observed for (a) a single mismatched dangling end, (b) an entirely mismatched dangling end, (c) a half-truncated dangling end with a mismatch, or (d) a mismatch with half-mismatched dangling end.

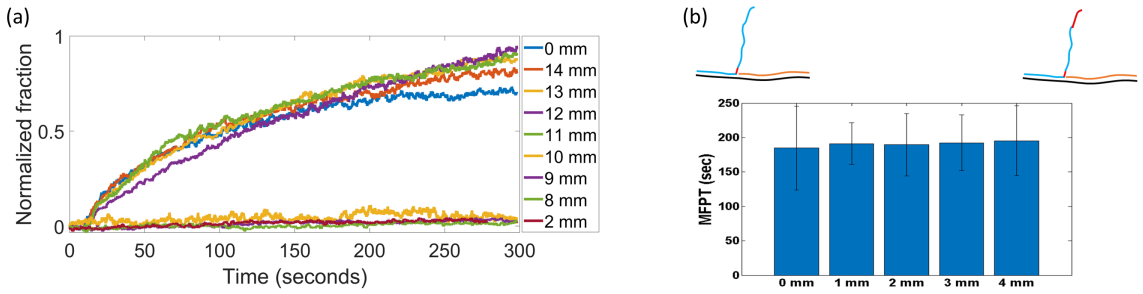


Figure A.2: Testing end invasion. All 14nt invaders had a first position mismatch. The number of consecutive toehold distal mismatches were varied. (a) Raw molecule counts were normalized and charted over time. The legend indicates the most proximal position for the consecutive mismatches. For up to 4 consecutive mismatches, strand displacement behaved identically to the invader with only a first position mismatch. Beyond that, no displacement was observed. (b) The abscissa marks the number of consecutive mismatches. The MFPT was extracted, and the experiment was repeated in triplicate. No difference in MFPT was observed for fewer than 5 consecutive mismatches.

half-truncated branch migration region, and half-mismatched branch migration region. We observed no strand displacement for any of these variations. These indicate the a mostly matching invader is necessary for strand displacement to proceed.

We considered the possibility that the invader halted at the first mismatch could continue displacement by invading at the end most distal to the toehold. Such a process would be similar to a zero-toehold invasion which is exponentially slower than strand displacement with a toehold [25]. We measured the MFPT of several invader constructs with consecutive mismatches occurring at the toehold distal end (see Figure A.2). For invaders with up to 4 consecutive distal mismatches, we found that strand displacement MFPT was identical to that of an invader with no distal mismatches. For 5 or more, there was a drastic change where no observed displacement occurred. Finally, we observed the behavior of an invader with 2 consecutive mismatches at the toehold proximal end and found no observable displacement occurred (see Figure A.3). Taken together, these results point to the possibility of alternative pathways of strand displacement in the presence of a mismatch.

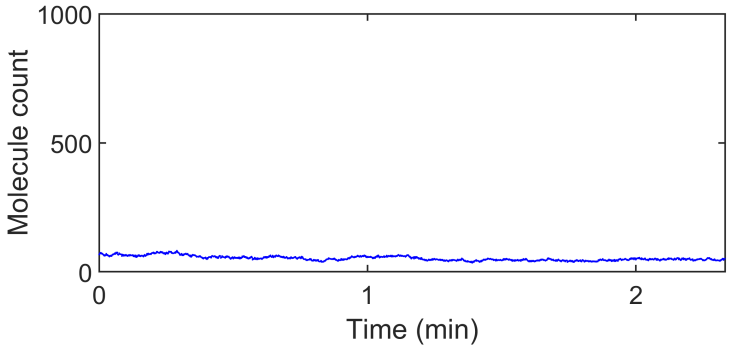


Figure A.3: 2 consecutive mismatches. No displacement was observed for invaders with consecutive mismatches in the first two toehold proximal base pairs.

REFERENCES

[1] J. D. Watson, F. H. Crick, *et al.*, "Molecular structure of nucleic acids," *Nature*, vol. 171, no. 4356, pp. 737–738, 1953.

[2] E. Schrodinger and Lewin, *What is life?* University Press, 1967.

[3] W. D. Wright, S. S. Shah, and W.-D. Heyer, "Homologous recombination and the repair of dna double-strand breaks," *Journal of Biological Chemistry*, jbc–TM118, 2018.

[4] E. S. Andersen, M. Dong, M. M. Nielsen, K. Jahn, R. Subramani, W. Mamdouh, M. M. Golas, B. Sander, H. Stark, C. L. Oliveira, *et al.*, "Self-assembly of a nanoscale dna box with a controllable lid," *Nature*, vol. 459, no. 7243, pp. 73–76, 2009.

[5] A. E. Marras, L. Zhou, H.-J. Su, and C. E. Castro, "Programmable motion of dna origami mechanisms," *PNAS*, vol. 112, no. 3, pp. 713–718, 2015.

[6] W. Rogers and V. Manoharan, "Programming colloidal phase transitions with dna strand displacement," *Bull. Am. Phys. Soc.*, vol. 59, 2014.

[7] Z. Zhang, D. Zeng, H. Ma, G. Feng, J. Hu, L. He, C. Li, and C. Fan, "A dna-origami chip platform for label-free snp genotyping using toehold-mediated strand displacement," *Small*, vol. 6, no. 17, pp. 1854–1858, 2010.

[8] L. Zhou, A. E. Marras, H.-J. Su, and C. E. Castro, "Direct design of an energy landscape with bistable dna origami mechanisms," *Nano Lett.*, vol. 15, no. 3, pp. 1815–1821, 2015.

[9] D. Y. Zhang, S. X. Chen, and P. Yin, "Optimizing the specificity of nucleic acid hybridization," *Nature chemistry*, vol. 4, no. 3, pp. 208–214, 2012.

[10] Z. Zhang, C. Hejesen, M. B. Kjelstrup, V. Birkedal, and K. V. Gothelf, "A dna-mediated homogeneous binding assay for proteins and small molecules," *Journal of the American Chemical Society*, vol. 136, no. 31, pp. 11 115–11 120, 2014.

[11] J. Bath and A. J. Turberfield, "Dna nanomachines," *Nature nanotechnology*, vol. 2, no. 5, pp. 275–284, 2007.

[12] W. B. Sherman and N. C. Seeman, "A precisely controlled dna biped walking device," *Nano Letters*, vol. 4, no. 7, pp. 1203–1207, 2004.

[13] J.-S. Shin and N. A. Pierce, "A synthetic dna walker for molecular transport," *Journal of the American Chemical Society*, vol. 126, no. 35, pp. 10 834–10 835, 2004.

[14] F. C. Simmel, "Processive motion of bipedal dna walkers," *ChemPhysChem*, vol. 10, no. 15, pp. 2593–2597, 2009.

[15] P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield, and J. H. Reif, "A unidirectional dna walker that moves autonomously along a track," *Angewandte Chemie International Edition*, vol. 43, no. 37, pp. 4906–4911, 2004.

[16] X. He, Z. Li, M. Chen, and N. Ma, "Dna-programmed dynamic assembly of quantum dots for molecular computation," *Angew. Chem. Int. Ed*, vol. 126, no. 52, pp. 14 675–14 678, 2014.

[17] X. Li, Z. Wang, W. Lu, Z. Chen, Y. Wang, and X. Shi, "A spiking neural system based on dna strand displacement," *J. Comput. Theor. Nanosci.*, vol. 12, no. 2, pp. 298–304, 2015.

[18] Y. Liu, B. Dong, Z. Wu, W. Fang, G. Zhou, A. Shen, X. Zhou, and J. Hu, "Toehold-mediated dna logic gates based on host–guest dna-gnps," *Chem. Commun.*, vol. 50, no. 81, pp. 12 026–12 029, 2014.

[19] L. Qian and E. Winfree, "Scaling up digital circuit computation with dna strand displacement cascades," *Science*, vol. 332, no. 6034, pp. 1196–1201, 2011.

[20] Y. Wang, G. Tian, H. Hou, M. Ye, and G. Cui, "Simple logic computation based on the dna strand displacement," *J. Comput. Theor. Nanosci.*, vol. 11, no. 9, pp. 1975–1982, 2014.

[21] C. Wu, S. Wan, W. Hou, L. Zhang, J. Xu, C. Cui, Y. Wang, J. Hu, and W. Tan, "A survey of advancements in nucleic acid-based logic gates and computing for applications in biotechnology and biomedicine," *Chem. Commun.*, 2015.

[22] E. Vischer and E. Chargaff, "The separation and quantitative estimation of purines and pyrimidines in minute amounts," *Journal of Biological Chemistry*, vol. 176, no. 2, pp. 703–714, 1948.

[23] L. Bao, X. Zhang, Y.-Z. Shi, Y.-Y. Wu, and Z.-J. Tan, "Understanding the relative flexibility of rna and dna duplexes: Stretching and twist-stretch coupling," *Biophysical journal*, vol. 112, no. 6, pp. 1094–1104, 2017.

[24] D. Y. Zhang and E. Winfree, "Control of dna strand displacement kinetics using toehold exchange," *J. Am. Chem. Soc.*, vol. 131, no. 47, pp. 17 303–17 314, 2009.

[25] N. Srinivas, T. E. Ouldridge, P. Šulc, J. M. Schaeffer, B. Yurke, A. A. Louis, J. P. Doye, and E. Winfree, "On the biophysics and kinetics of toehold-mediated dna strand displacement," *Nucleic Acids Res.*, vol. 41, no. 22, pp. 10 641–10 658, 2013.

[26] R. A. Epstein, *The theory of gambling and statistical logic*. Academic Press, 2012.

[27] R. Carlson. (2014). Time for new dna synthesis and sequencing cost curves.

[28] A. J. Thubagere, W. Li, R. F. Johnson, Z. Chen, S. Doroudi, Y. L. Lee, G. Izatt, S. Wittman, N. Srinivas, D. Woods, *et al.*, "A cargo-sorting dna robot," *Science*, vol. 357, no. 6356, eaan6558, 2017.

[29] F. Dannenberg, M. Kwiatkowska, C. Thachuk, and A. J. Turberfield, "Dna walker circuits: Computational potential, design, and verification," *Natural Computing*, vol. 14, no. 2, pp. 195–211, 2015.

[30] J. Pan, F. Li, T.-G. Cha, H. Chen, and J. H. Choi, "Recent progress on dna based walkers," *Curr. Opin. Biotechnol.*, vol. 34, pp. 56–64, 2015.

[31] C. Jung, P. B. Allen, and A. D. Ellington, "A simple, cleated dna walker that hangs on to surfaces," *ACS nano*, vol. 11, no. 8, pp. 8047–8054, 2017.

[32] Y. Ke, T. Meyer, W. M. Shih, and G. Bellot, "Regulation at a distance of biomolecular interactions using a dna origami nanoactuator," *Nature Communications*, vol. 7, 2016.

[33] D. Y. Zhang and G. Seelig, "Dynamic dna nanotechnology using strand-displacement reactions," *Nat. Chem.*, vol. 3, no. 2, pp. 103–113, 2011.

[34] R. Ranjbar and M. S. Hafezi-Moghadam, "Design and construction of a dna origami drug delivery system based on mpt64 antibody aptamer for tuberculosis treatment," *Electronic physician*, vol. 8, no. 2, p. 1857, 2016.

[35] Z. Zhao, Y. Liu, and H. Yan, "Organizing dna origami tiles into larger structures using preformed scaffold frames," *Nano letters*, vol. 11, no. 7, pp. 2997–3002, 2011.

[36] P. Chidchob, T. G. Edwardson, C. J. Serpell, and H. F. Sleiman, "Synergy of two assembly languages in dna nanostructures: Self-assembly of sequence-defined polymers on dna cages," *Journal of the American Chemical Society*, 2015.

[37] A. K. George and H. Singh, "Dna strand displacement-based logic inverter gate design," *Micro & Nano Letters*, vol. 12, no. 9, pp. 611–614, 2017.

[38] D. Y. Zhang, A. J. Turberfield, B. Yurke, and E. Winfree, "Engineering entropy-driven reactions and networks catalyzed by dna," *Science*, vol. 318, no. 5853, pp. 1121–1125, 2007.

[39] X. J. Chen, "Mechanism of homologous recombination and implications for aging-related deletions in mitochondrial dna," *Microbiology and Molecular Biology Reviews*, vol. 77, no. 3, pp. 476–496, 2013.

[40] F. J. Mojica, J. García-Martínez, E. Soria, *et al.*, "Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements," *Journal of molecular evolution*, vol. 60, no. 2, pp. 174–182, 2005.

[41] E. Deltcheva, K. Chylinski, C. M. Sharma, K. Gonzales, Y. Chao, Z. A. Pirzada, M. R. Eckert, J. Vogel, and E. Charpentier, "Crispr rna maturation by trans-encoded small rna and host factor rnase iii," *Nature*, vol. 471, no. 7340, p. 602, 2011.

[42] D. Singh, S. H. Sternberg, J. Fei, J. A. Doudna, and T. Ha, "Real-time observation of dna recognition and rejection by the rna-guided endonuclease cas9," *Nature communications*, vol. 7, p. 12 778, 2016.

[43] M. KIreeva, C. Trang, G. Matevosyan, J. Turek-Herman, V. Chasov, L. Lubkowska, and M. Kashlev, "Rna–dna and dna–dna base-pairing at the upstream edge of the transcription bubble regulate translocation of rna polymerase and transcription rate," *Nucleic acids research*, vol. 46, no. 11, pp. 5764–5775, 2018.

[44] B. Yurke and A. P. Mills Jr, "Using dna to power nanostructures," *Genetic Programming and Evolvable Machines*, vol. 4, no. 2, pp. 111–122, 2003.

[45] P. Šulc, F. Romano, T. E. Ouldridge, L. Rovigatti, J. P. Doye, and A. A. Louis, "Sequence-dependent thermodynamics of a coarse-grained dna model," *The Journal of chemical physics*, vol. 137, no. 13, p. 135 101, 2012.

[46] R. R. Machinek, T. E. Ouldridge, N. E. Haley, J. Bath, and A. J. Turberfield, "Programmable energy landscapes for kinetic control of dna strand displacement," *Nat. Commun.*, vol. 5, 2014.

[47] G. Mendel, "Versuche über pflanzenhybriden," *Verhandlungen des naturforschenden Vereines in Brunn 4: 3*, vol. 44, 1866.

[48] E. Lander. (2004). 7.012.

[49] E. Neher, B. Sakmann, and J. H. Steinbach, "The extracellular patch clamp: A method for resolving currents through individual open channels in biological membranes," *Pflügers Archiv*, vol. 375, no. 2, pp. 219–228, 1978.

[50] K. C. Neuman and A. Nagy, "Single-molecule force spectroscopy: Optical tweezers, magnetic tweezers and atomic force microscopy," *Nature methods*, vol. 5, no. 6, p. 491, 2008.

[51] T. Ha, "Probing nature's nanomachines one molecule at a time.," *Biophysical journal*, vol. 110, no. 5, p. 1004, 2016.

[52] W.-Y. Jiang and S.-Y. Ran, "Two-stage dna compaction induced by silver ions suggests a cooperative binding mechanism," *The Journal of chemical physics*, vol. 148, no. 20, p. 205 102, 2018.

[53] B. Gollnick, C. Carrasco, F. Zuttion, N. S. Gilhooly, M. S. Dillingham, and F. Moreno-Herrero, "Probing dna helicase kinetics with temperature-controlled magnetic tweezers," *small*, vol. 11, no. 11, pp. 1273–1284, 2015.

[54] M. Rutkauskas, T. Sinkunas, I. Songailiene, M. S. Tikhomirova, V. Siksnys, and R. Seidel, "Directional r-loop formation by the crispr-cas surveillance complex cascade provides efficient off-target site rejection," *Cell reports*, vol. 10, no. 9, pp. 1534–1543, 2015.

[55] U. Bockelmann, P. Thomen, B Essevaz-Roulet, V. Viasnoff, and F. Heslot, "Unzipping dna with optical tweezers: High sequence sensitivity and force flips," *Biophysical journal*, vol. 82, no. 3, pp. 1537–1553, 2002.

[56] M. D. Wang, H. Yin, R. Landick, J. Gelles, and S. M. Block, "Stretching dna with optical tweezers," *Biophysical journal*, vol. 72, no. 3, pp. 1335–1346, 1997.

[57] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule fret," *Nature methods*, vol. 5, no. 6, p. 507, 2008.

[58] S. Vernick, S. M. Trocchia, S. B. Warren, E. F. Young, D. Bouilly, R. L. Gonzalez, C. Nuckolls, and K. L. Shepard, "Electrostatic melting in a single-molecule field-effect transistor with applications in genomic identification," *Nature communications*, vol. 8, p. 15 450, 2017.

[59] T. Scientific. (2018). Cy5 spectra.

[60] H. H. Jaffe and A. L. Miller, "The fates of electronic excitation energy," *Journal of Chemical Education*, vol. 43, no. 9, p. 469, 1966.

[61] L. Guo and F. Gai, "Simple method to enhance the photostability of the fluorescence reporter r6g for prolonged single-molecule studies," *The Journal of Physical Chemistry A*, vol. 117, no. 29, pp. 6164–6170, 2013.

[62] T. Ha and P. Tinnefeld, "Photophysics of fluorescent probes for single-molecule biophysics and super-resolution imaging," *Annual review of physical chemistry*, vol. 63, pp. 595–617, 2012.

[63] B. Schuler and W. A. Eaton, "Protein folding studied by single-molecule fret," *Current opinion in structural biology*, vol. 18, no. 1, pp. 16–26, 2008.

[64] S. H. Sternberg, B. LaFrance, M. Kaplan, and J. A. Doudna, "Conformational control of dna target cleavage by crispr–cas9," *Nature*, vol. 527, no. 7576, p. 110, 2015.

[65] T. T. Le and H. D. Kim, "Probing the elastic limit of dna bending," *Nucleic Acids Res.*, vol. 42, no. 16, pp. 10 786–10 794, 2014.

[66] B. Masters, "Paths to förster's resonance energy transfer (fret) theory," *The European Physical Journal H*, vol. 39, no. 1, pp. 87–139, 2014.

[67] A. Clayborne and V. Morris. (2017). Physical chemistry laboratory.

[68] J. Autschbach, "Why the particle-in-a-box model works well for cyanine dyes but not for conjugated polyenes," *Journal of chemical education*, vol. 84, no. 11, p. 1840, 2007.

[69] A. Zangwill, *Modern electrodynamics*. Cambridge University Press, 2013.

[70] C. Joo, H. Balci, Y. Ishitsuka, C. Buranachai, and T. Ha, "Advances in single-molecule fluorescence methods for molecular biology," *Annu. Rev. Biochem.*, vol. 77, pp. 51–76, 2008.

[71] H. Zhang and P. Guo, "Single molecule photobleaching (smpb) technology for counting of rna, dna, protein and other molecules in nanoparticles and biological complexes by tirf instrumentation," *Methods*, vol. 67, no. 2, pp. 169–176, 2014.

[72] L. J. Friedman, J. Chung, and J. Gelles, "Viewing dynamic assembly of molecular complexes by multi-wavelength single-molecule fluorescence," *Biophysical journal*, vol. 91, no. 3, pp. 1023–1031, 2006.

[73] J. P. Zinter and M. J. Levene, "Maximizing fluorescence collection efficiency in multiphoton microscopy," *Optics express*, vol. 19, no. 16, pp. 15 348–15 362, 2011.

[74] Y. Fang, "Total internal reflection fluorescence quantification of receptor pharmacology," *Biosensors*, vol. 5, no. 2, pp. 223–240, 2015.

[75] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri, "Imagej2: Imagej for the next generation of scientific image data," *BMC bioinformatics*, vol. 18, no. 1, p. 529, 2017.

[76] M. Ovesnỳ, P. Křížek, J. Borkovec, Z. Švindrych, and G. M. Hagen, "Thunderstorm: A comprehensive imagej plug-in for palm and storm data analysis and super-resolution imaging," *Bioinformatics*, vol. 30, no. 16, pp. 2389–2390, 2014.

[77] T. E. Tomov, R. Tsukanov, M. Liber, R. Masoud, N. Plavner, and E. Nir, "Rational design of dna motors: Fuel optimization through single-molecule fluorescence," *J. Am. Chem. Soc.*, vol. 135, no. 32, pp. 11 935–11 941, 2013.

[78] W. L. Hwang, S. Deindl, B. T. Harada, and X. Zhuang, "Histone h4 tail mediates allosteric regulation of nucleosome remodelling by linker dna," *Nature*, vol. 512, no. 7513, p. 213, 2014.

[79] T. T. Ngo, J. Yoo, Q. Dai, Q. Zhang, C. He, A. Aksimentiev, and T. Ha, "Effects of cytosine modifications on dna flexibility and nucleosome mechanical stability," *Nature communications*, vol. 7, p. 10 813, 2016.

[80] T. D. Christian, L. J. Romano, and D. Rueda, "Single-molecule measurements of synthesis by dna polymerase with base-pair resolution," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21 109–21 114, 2009.

[81] R. Lamichhane, A. Solem, W. Black, and D. Rueda, "Single-molecule fret of protein–nucleic acid and protein–protein complexes: Surface passivation and immobilization," *Methods*, vol. 52, no. 2, pp. 192–200, 2010.

[82] F. Rashid, P. D. Harris, M. S. Zaher, M. A. Sobhy, L. I. Joudeh, C. Yan, H. Piwonski, S. E. Tsutakawa, I. Ivanov, J. A. Tainer, *et al.*, "Single-molecule fret unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1," *Elife*, vol. 6, e21884, 2017.

[83] J. L. Alejo and S. C. Blanchard, "Miscoding-induced stalling of substrate translocation on the bacterial ribosome," *Proceedings of the National Academy of Sciences*, vol. 114, no. 41, E8603–E8610, 2017.

[84] S. Arenz, M. F. Juette, M. Graf, F. Nguyen, P. Huter, Y. S. Polikanov, S. C. Blanchard, and D. N. Wilson, "Structures of the orthosomycin antibiotics avilamycin and evernimicin in complex with the bacterial 70s ribosome," *Proceedings of the National Academy of Sciences*, vol. 113, no. 27, pp. 7527–7532, 2016.

[85] A. I. Cocozaki, R. B. Altman, J. Huang, E. T. Buurman, S. L. Kazmirski, P. Doig, D. B. Prince, S. C. Blanchard, J. H. Cate, and A. D. Ferguson, "Resistance mutations generate divergent antibiotic susceptibility profiles against translation inhibitors," *Proceedings of the National Academy of Sciences*, vol. 113, no. 29, pp. 8188–8193, 2016.

[86] A. Ferguson, L. Wang, R. B. Altman, D. S. Terry, M. F. Juette, B. J. Burnett, J. L. Alejo, R. A. Dass, M. M. Parks, C. T. Vincent, *et al.*, "Functional dynamics within the human ribosome regulate the rate of active protein synthesis," *Molecular cell*, vol. 60, no. 3, pp. 475–486, 2015.

[87] M. R. Wasserman, J. L. Alejo, R. B. Altman, and S. C. Blanchard, "Multiperspective smfret reveals rate-determining late intermediates of ribosomal translocation," *Nature structural & molecular biology*, vol. 23, no. 4, p. 333, 2016.

[88] N. Vušurović, R. B. Altman, D. S. Terry, R. Micura, and S. C. Blanchard, "Pseudoknot formation seeds the twister ribozyme cleavage reaction coordinate," *Journal of the American Chemical Society*, vol. 139, no. 24, pp. 8186–8193, 2017.

[89] C. Manz, A. Y. Kobitski, A. Samanta, B. G. Keller, A. Jäschke, and G. U. Nienhaus, "Single-molecule fret reveals the energy landscape of the full-length sam-i riboswitch," *Nature chemical biology*, vol. 13, no. 11, p. 1172, 2017.

[90] V. C. DeRocco, L. E. Sass, R. Qiu, K. R. Weninger, and D. A. Erie, "Dynamics of muts–mismatched dna complexes are predictive of their repair phenotypes," *Biochemistry*, vol. 53, no. 12, pp. 2043–2052, 2014.

[91] D. A. Erie and K. R. Weninger, "Single molecule studies of dna mismatch repair," *DNA repair*, vol. 20, pp. 71–81, 2014.

[92] R. Qiu, M. Sakato, E. J. Sacho, H. Wilkins, X. Zhang, P. Modrich, M. M. Hingorani, D. A. Erie, and K. R. Weninger, "Mutl traps muts at a dna mismatch," *Proceedings of the National Academy of Sciences*, vol. 112, no. 35, pp. 10 914–10 919, 2015.

[93] J. S. Chen, Y. S. Dagdas, B. P. Kleinstiver, M. M. Welch, A. A. Sousa, L. B. Harrington, S. H. Sternberg, J. K. Joung, A. Yildiz, and J. A. Doudna, "Enhanced proofreading governs crispr–cas9 targeting accuracy," *Nature*, vol. 550, no. 7676, p. 407, 2017.

[94] D. B. Broadwater and H. D. Kim, "The effect of basepair mismatch on dna strand displacement," *Biophysical journal*, vol. 110, no. 7, pp. 1476–1484, 2016.

[95] J. Jeong, T. T. Le, and H. D. Kim, "Single-molecule fluorescence studies on dna looping," *Methods*, vol. 105, pp. 34–43, 2016.

[96] B. Hua, K. Y. Han, R. Zhou, H. Kim, X. Shi, S. C. Abeysirigunawardena, A. Jain, D. Singh, V. Aggarwal, S. A. Woodson, *et al.*, "An improved surface passivation method for single-molecule studies," *nature methods*, vol. 11, no. 12, p. 1233, 2014.

[97] R. Holliday, "A mechanism for gene conversion in fungi," *Genetical Research*, vol. 5, no. 02, pp. 282–304, 1964.

[98] S. P. Jackson, "Sensing and repairing dna double-strand breaks," *Carcinogenesis*, vol. 23, no. 5, pp. 687–696, 2002.

[99] J. R. Chapman, M. R. Taylor, and S. J. Boulton, "Playing the end game: Dna double-strand break repair pathway choice," *Molecular cell*, vol. 47, no. 4, pp. 497–510, 2012.

[100] Q. Li, G. Luan, Q. Guo, and J. Liang, "A new class of homogeneous nucleic acid probes based on specific displacement hybridization," *Nucleic Acids Res.*, vol. 30, no. 2, e5–e5, 2002.

[101] A. A. Green, P. A. Silver, J. J. Collins, and P. Yin, "Toehold switches: De-novo-designed regulators of gene expression," *Cell*, vol. 159, no. 4, pp. 925–939, 2014.

[102] Y. S. Jiang, S. Bhadra, B. Li, Y. R. Wu, J. N. Milligan, and A. D. Ellington, "Robust strand exchange reactions for the sequence-specific, real-time detection of nucleic acid amplicons," *Anal. Chem.*, vol. 87, no. 6, pp. 3314–3320, 2015.

[103] R. Schulman and D. Doty, "Designing ordered nucleic acid self-assembly processes," *Curr. Opin. Struct. Biol.*, vol. 31, pp. 57–63, 2015.

[104] M. Liber, T. E. Tomov, R. Tsukanov, Y. Berger, and E. Nir, "A bipedal dna motor that travels back and forth between two dna origami tiles," *Small*, vol. 11, no. 5, pp. 568–575, 2015.

[105] P. B. Landon, J. Lee, M. T. Hwang, A. H. Mo, C. Zhang, A. Neuberger, B. Meckes, J. J. Gutierrez, G. Glinsky, and R. Lal, "Energetically biased dna motor containing a thermodynamically stable partial strand displacement state," *Langmuir*, vol. 30, no. 46, pp. 14 073–14 078, 2014.

[106] R. Masoud, R. Tsukanov, T. E. Tomov, N. Plavner, M. Liber, and E. Nir, "Studying the structural dynamics of bipedal dna motors with single-molecule fluorescence spectroscopy," *ACS Nano*, vol. 6, no. 7, pp. 6272–6283, 2012.

[107] W. Lee, P. H. von Hippel, and A. H. Marcus, "Internally labeled cy3/cy5 dna constructs show greatly enhanced photo-stability in single-molecule fret experiments," *Nucleic Acids Res.*, gku199, 2014.

[108] T. T. Le and H. D. Kim, "Studying dna looping by single-molecule fret," *JoVE (Journal of Visualized Experiments)*, no. 88, e51667–e51667, 2014.

[109] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, "An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments," *Biophys. J.*, vol. 94, no. 5, pp. 1826–1835, 2008.

[110] S. K. Kim, "Mean first passage time for a random walker and its application to chemical kinetics," *J. Chem. Phys.*, vol. 28, no. 6, pp. 1057–1067, 1958.

[111] K. P. N. Murthy and K. W. Kehr, "Mean first-passage time of random walks on a random lattice," *Phys. Rev. A*, vol. 40, pp. 2082–2087, 4 1989.

[112] A. Mentes, A. M. Florescu, E. Brunk, J. Wereszczynski, M. Joyeux, and I. Andricioaei, "Free-energy landscape and characteristic forces for the initiation of dna unzipping," *Biophys. J.*, vol. 108, no. 7, pp. 1727–1738, 2015.

[113] G. Bel, B. Munsky, and I. Nemenman, "The simplicity of completion time distributions for common complex biochemical processes," *Physical biology*, vol. 7, no. 1, p. 016 003, 2010.

[114] V. Anshelevich, A. Vologodskii, A. Lukashin, and M. Frank-Kamenetskii, "Slow relaxational processes in the melting of linear biopolymers: A theory and its application to nucleic acids," *Biopolymers*, vol. 23, no. 1, pp. 39–58, 1984.

[115] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Res.*, vol. 31, no. 13, pp. 3406–3415, 2003.

[116] P. A. Pury and M. O. Cáceres, "Mean first-passage and residence times of random walks on asymmetric disordered chains," *Journal of Physics A: Mathematical and General*, vol. 36, no. 11, p. 2695, 2003.

[117] G Lakatos, T Chou, B Bergersen, and G Patey, "First passage times of pulling-assisted dna unzipping," *arXiv preprint cond-mat/0406006*, 2004.

[118] Y. Gao, L. K. Wolf, and R. M. Georgiadis, "Secondary structure effects on dna hybridization kinetics: A solution versus surface comparison," *Nucleic Acids Res.*, vol. 34, no. 11, pp. 3370–3377, 2006.

[119] I. I. Cisse, H. Kim, and T. Ha, "A rule of seven in watson-crick base-pairing of mismatched sequences," *Nat. Struct. Mol. Biol.*, vol. 19, no. 6, pp. 623–627, 2012.

[120] S. Cocco, R. Monasson, and J. F. Marko, "Force and kinetic barriers to unzipping of the dna double helix," *PNAS*, vol. 98, no. 15, pp. 8608–8613, 2001.

[121] N. G. Van Kampen, *Stochastic processes in physics and chemistry*. Elsevier, 1992, vol. 1.

[122] D. L. Floyd, S. C. Harrison, and A. M. Van Oijen, "Analysis of kinetic intermediates in single-particle dwell-time distributions," *Biophysical journal*, vol. 99, no. 2, pp. 360–366, 2010.

[123] M. D. Betterton and F. Jülicher, "Opening of nucleic-acid double strands by helicases: Active versus passive opening," *Phys. Rev. E*, vol. 71, p. 011 904, 1 2005.

[124] I. G. Panyutin and P. Hsieh, "The kinetics of spontaneous dna branch migration," *PNAS*, vol. 91, no. 6, pp. 2021–2025, 1994.

[125] J. T. Waters and H. D. Kim, "Calculation of a fluctuating entropic force by phase space sampling," *Physical Review E*, vol. 92, no. 1, p. 013 308, 2015.

[126] T. Paramanathan, D. Reeves, L. J. Friedman, J. Kondev, and J. Gelles, "A general mechanism for competitor-induced dissociation of molecular complexes," *Nat. Commun.*, vol. 5, 2014.

[127] P. J. Sanstead and A. Tokmakoff, "Direct observation of activated kinetics and downhill dynamics in dna dehybridization," *The Journal of Physical Chemistry B*, vol. 122, no. 12, pp. 3088–3100, 2018.

[128] S. A. McKinney, A.-C. Déclais, D. M. Lilley, and T. Ha, "Structural dynamics of individual holliday junctions," *Nature Structural and Molecular Biology*, vol. 10, no. 2, p. 93, 2003.

[129] D. V. Bugreev, O. M. Mazina, and A. V. Mazin, "Rad54 protein promotes branch migration of holliday junctions," *Nature*, vol. 442, no. 7102, p. 590, 2006.

[130] C. M. Radding, K. L. Beattie, W. K. Holloman, and R. C. Wiegand, "Uptake of homologous single-stranded fragments by superhelical dna: Iv. branch migration," *Journal of molecular biology*, vol. 116, no. 4, pp. 825–839, 1977.

[131] C. Green and C. Tibbetts, "Reassociation rate limited displacement of dna strands by branch migration," *Nucleic Acids Research*, vol. 9, no. 8, pp. 1905–1918, 1981.

[132] P. Šulc, T. E. Ouldridge, F. Romano, J. P. Doye, and A. A. Louis, "Modelling toehold-mediated rna strand displacement," *Biophysical journal*, vol. 108, no. 5, pp. 1238–1247, 2015.

[133] A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman, "Advanced methods of microscope control using μmanager software," *Journal of biological methods*, vol. 1, no. 2, 2014.

[134] K. J. Breslauer, R. Frank, H. Blöcker, and L. A. Marky, "Predicting dna duplex stability from the base sequence," *Proceedings of the National Academy of Sciences*, vol. 83, no. 11, pp. 3746–3750, 1986.

[135] J. SantaLucia, H. T. Allawi, and P. A. Seneviratne, "Improved nearest-neighbor parameters for predicting dna duplex stability," *Biochemistry*, vol. 35, no. 11, pp. 3555–3562, 1996.

[136] J. SantaLucia, "A unified view of polymer, dumbbell, and oligonucleotide dna nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, pp. 1460–1465, 1998.

[137] N. Sugimoto, S.-i. Nakano, M. Yoneyama, and K.-i. Honda, "Improved thermodynamic parameters and helix initiation factor to predict stability of dna duplexes," *Nucleic acids research*, vol. 24, no. 22, pp. 4501–4505, 1996.

[138] K. Tawa and W. Knoll, "Mismatching base-pair dependence of the kinetics of dna–dna hybridization studied by surface plasmon fluorescence spectroscopy," *Nucleic acids research*, vol. 32, no. 8, pp. 2372–2377, 2004.

[139] J. Fiche, A Buhot, R Calemczuk, and T Livache, "Temperature effects on dna chip experiments from surface plasmon resonance imaging: Isotherms and melting curves," *Biophysical journal*, vol. 92, no. 3, pp. 935–946, 2007.

[140] A. Gunnarsson, P. Jönsson, V. P. Zhdanov, and F. Höök, "Kinetic and thermodynamic characterization of single-mismatch discrimination using single-molecule imaging," *Nucleic acids research*, vol. 37, no. 14, e99–e99, 2009.

[141] H. Mundigala, J. B. Michaux, A. L. Feig, E. Ennifar, and D. Rueda, "Hiv-1 dis stem loop forms an obligatory bent kissing intermediate in the dimerization pathway," *Nucleic acids research*, vol. 42, no. 11, pp. 7281–7289, 2014.

[142] J. Kim, S. Doose, H. Neuweiler, and M. Sauer, "The initial step of dna hairpin folding: A kinetic analysis using fluorescence correlation spectroscopy," *Nucleic acids research*, vol. 34, no. 9, pp. 2516–2527, 2006.

[143] P. L. Biancaniello, A. J. Kim, and J. C. Crocker, "Long-time stretched exponential kinetics in single dna duplex dissociation," *Biophysical journal*, vol. 94, no. 3, pp. 891–896, 2008.

[144] E. D. Holmstrom, N. F. Dupuis, and D. J. Nesbitt, "Pulsed ir heating studies of single-molecule dna duplex dissociation kinetics and thermodynamics," *Biophysical journal*, vol. 106, no. 1, pp. 220–231, 2014.

[145] S Ikuta, K Takagi, R. B. Wallace, and K Itakura, "Dissociation kinetics of 19 base paired oligonucleotide-dna duplexes containing different single mismatched base pairs," *Nucleic Acids Research*, vol. 15, no. 2, pp. 797–811, 1987.

[146] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, "Unzipping kinetics of double-stranded dna in a nanopore," *Physical Review Letters*, vol. 90, no. 23, p. 238 101, 2003.

[147] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante, "Reversible unfolding of single rna molecules by mechanical force," *Science*, vol. 292, no. 5517, pp. 733–737, 2001.

[148] J. Svitel, H. Boukari, D. Van Ryk, R. C. Willson, and P. Schuck, "Probing the functional heterogeneity of surface binding sites by analysis of experimental binding traces and the effect of mass transport limitation," *Biophysical journal*, vol. 92, no. 5, pp. 1742–1758, 2007.

[149] V. Tejedor, M. Schad, O. Bénichou, R. Voituriez, and R. Metzler, "Encounter distribution of two random walkers on a finite one-dimensional interval," *Journal of Physics A: Mathematical and Theoretical*, vol. 44, no. 39, p. 395 005, 2011.

[150] S. Bommarito, N. Peyret, and J. SantaLucia Jr, "Thermodynamic parameters for dna sequences with dangling ends," *Nucleic acids research*, vol. 28, no. 9, pp. 1929–1934, 2000.

[151] T. Ohmichi, S.-i. Nakano, D. Miyoshi, and N. Sugimoto, "Long rna dangling end has large energetic contribution to duplex stability," *Journal of the American Chemical Society*, vol. 124, no. 35, pp. 10 367–10 372, 2002.

[152] P. V. Riccelli, K. E. Mandell, and A. S. Benight, "Melting studies of dangling-ended dna hairpins: Effects of end length, loop sequence and biotinylation of loop bases," *Nucleic acids research*, vol. 30, no. 18, pp. 4088–4093, 2002.

[153] S. Mohan, C. Hsiao, H. VanDeusen, R. Gallagher, E. Krohn, B. Kalahar, R. M. Wartell, and L. D. Williams, "Mechanism of rna double helix-propagation at atomic resolution," *The Journal of Physical Chemistry B*, vol. 113, no. 9, pp. 2614–2623, 2009.

[154] I. Anosova, E. A. Kowal, M. R. Dunn, J. C. Chaput, W. D. Van Horn, and M. Egli, "The structural diversity of artificial genetic polymers," *Nucleic acids research*, vol. 44, no. 3, pp. 1007–1021, 2016.

[155] M. A. Spies and R. L. Schowen, "The trapping of a spontaneously "flipped-out" base from double helical nucleic acids by host- guest complexation with β-cyclodextrin: The intrinsic base-flipping rate constant for dna and rna," *Journal of the American Chemical Society*, vol. 124, no. 47, pp. 14 049–14 053, 2002.

[156] S. A. Sevier and H. Levine, "Properties of gene expression and chromatin structure with mechanically regulated elongation," *Nucleic acids research*, 2018.

[157] Y. G. Sinai, "The limiting behavior of a one-dimensional random walk in a random medium," *Theory of Probability & Its Applications*, vol. 27, no. 2, pp. 256–268, 1983.

[158] S. Noskowicz and I Goldhirsch, "First-passage-time distribution in a random random walk," *Physical Review A*, vol. 42, no. 4, p. 2047, 1990.

[159] E. C. Hayden, "Is the $ 1,000 genome for real?" *Nature News*, 2014.

[160] P. Rajagopal and J. Feigon, "Triple-strand formation in the homopurine: Homopyrimidine dna oligonucleotides d (ga) 4 and d (tc) 4," *Nature*, vol. 339, no. 6226, p. 637, 1989.

[161] A. Dayn, G. M. Samadashwily, and S. M. Mirkin, "Intramolecular dna triplexes: Unusual sequence requirements and influence on dna polymerization," *Proceedings of the National Academy of Sciences*, vol. 89, no. 23, pp. 11 406–11 410, 1992.

VITA

Douglas William Bo Broadwater, Jr., was born in Columbus, Georgia, in 1987. He grew up in Phenix City, Alabama, where he attended Glenwood School, graduating in 2005. He attended Mercer University in Macon, Georgia, where he received a B.S.E. with a specialty in biomedical engineering and a B.A. in physics in 2010. After working for two years as an Electronics Engineer at Robins Air Force Base in Warner Robins, Georgia, he began graduate school in 2012 at the Georgia Institute of Technology in Atlanta, Georgia, where he earned his M.S. in physics in 2013, and conducted research under the advisement of Assoc. Prof. Harold Kim. He obtained his Ph.D. in physics in 2018.